A Multilevel Stochastic Collocation Method for Partial Differential Equations with Random Input Data

Vol. , pp.

A multilevel stochastic collocation method for partial differential equations with random input data

A. L. Teckentrup*, P. Jantsch†, C. G. Webster‡, and M. Gunzburger§

Abstract. Stochastic collocation methods for approximating the solution of partial differential equations with random input data (e.g., coefficients and forcing terms) suffer from the curse of dimensionality whereby increases in the stochastic dimension cause an explosion of the computational effort. We propose and analyze a multilevel version of the stochastic collocation method that, as is the case for multilevel Monte Carlo (MLMC) methods, uses hierarchies of spatial approximations to reduce the overall computational complexity. In addition, our proposed approach utilizes, for approximation in stochastic space, a sequence of multi-dimensional interpolants of increasing fidelity which can then be used for approximating statistics of the solution as well as for building high-order surrogates featuring faster convergence rates. A rigorous convergence and computational cost analysis of the new multilevel stochastic collocation method is provided in the case of elliptic equations, demonstrating its advantages compared to standard single-level stochastic collocation approximations as well as MLMC methods. Numerical results are provided that illustrate the theory and the effectiveness of the new multilevel method.

Key words. multilevel methods, stochastic collocation, PDEs with random input data, sparse grids, uncertainty quantification, finite element methods, multivariate polynomial approximation, hierarchical methods, high-dimensional approximation

AMS subject classifications. 65C20, 65C30, 65N30, 65N35, 65M75, 65T50, 65T60

1. Introduction.

Nowadays, mathematical modeling and computer simulations are used extensively in many scientific and engineering fields, usually with the goal of understanding or predicting the behavior of a system given its inputs such as the computational domain, model parameter values, and source terms. However, whether stemming from incomplete or inaccurate knowledge or from some inherent variability in the system, often these inputs may be subject to uncertainty. In order to correctly predict the behavior of the system, it is especially pertinent to understand and propagate the effect of the input uncertainty to the output of the simulation, i.e., to the solution of the mathematical model.

In this paper, we consider systems which are modeled by elliptic partial differential equations (PDEs) with random input data. We work under the *finite-dimensional noise assumption*, i.e., we assume that the random inputs are characterized by a finite-dimensional random vector. When enough information is available to completely characterize the randomness in the inputs, probability theory provides a natural setting for quantifying uncertainties. The object of our computations is the accurate calculation of solutions of stochastic elliptic PDEs

*Department of Scientific Computing, Florida State University, 400 Dirac Science Library, Tallahassee, FL 32306-4120 (ateckentrup@fsu.edu).

†Department of Mathematics, University of Tennessee, 1403 Circle Drive, Knoxville, TN 37916 (pjantsch@utk.edu).

‡Department of Computational and Applied Mathematics, Oak Ridge National Laboratory, One Bethel Valley Road, P.O. Box 2008, MS-6164, Oak Ridge, TN 37831-6164 (webstercg@ornl.gov).

§Department of Scientific Computing, Florida State University, 400 Dirac Science Library, Tallahassee, FL 32306-4120 (gunzburg@fsu.edu).

or statistics of some functional of the solution of the PDE. For instance, in addition to the solution itself, one might be interested in the expected value or variance of the solution in a given region of the computational domain.

A large number of methods have been developed for the numerical solution of PDEs with random inputs; see, e.g., [24] and the references cited therein. The most popular approach is the Monte Carlo (MC) method which involves random sampling of the input vector of random variables (also referred to as the *stochastic parameter space*) and the solution of the deterministic PDE at each of the sample points. In addition to the benefits of simple implementation and a natural decoupling of the stochastic and spatial degrees of freedom, MC methods feature a convergence rate that is independent of the dimension of the stochastic space. This makes it particularly attractive for high-dimensional problems. However, the convergence is in general very slow and, especially in case the stochastic space is only of moderate dimension and the solution of the PDE or a functional of interest is smooth with respect to the random parameters, better convergence rates can be achieved using more sophisticated methods.

Stochastic collocation (SC) methods [1, 31, 32] are similar to MC methods in the sense that they involve only the solution of a sequence of deterministic PDEs at given sample points in the stochastic space. However, rather than randomly chosen samples, SC methods use a deterministic grid of points at which one solves the corresponding deterministic PDE, and then builds an interpolant, either using global Lagrange-type polynomials [1, 31, 32] or even local hierarchical basis functions [23, 29]. For problems where the solution is a smooth function of the random input variables and the dimension of the stochastic space is moderate, SC methods have been shown to converge much faster than MC methods.

Unfortunately, for most problems, stochastic collocation methods suffer from the *curse of dimensionality*, a phrase that refers to the deterioration of the convergence rate and the explosion of computational effort as the dimension of the stochastic space increases. In this paper, we introduce a multilevel stochastic collocation (MLSC) approach for reducing the computational cost incurred by standard, i.e., single level, SC methods. Drawing inspiration from multigrid solvers for linear equations, the main idea behind multilevel methods is to utilize a hierarchical sequence of spatial approximations to the underlying PDE model that are then combined with stochastic discretizations in such a way as to minimize computational cost. Starting with the pioneering works [27] in the field of integral equations and [20] in the field of computational finance, the multilevel approach has been successfully applied to many applications of MC methods; see, e.g., [2, 8, 15, 21, 22, 28, 30].

The MLSC method proposed in this work is similar to the construction found in [4], where the authors propose a simple rule to balance the resolution of the spatial and stochastic discretizations, thereby reducing the total degrees of freedom. In contrast, our construction provides the flexibility of optimizing the interpolation operators used at each level of discretization to minimize computational cost. Our method is also similar to the multilevel quadrature approximations of moments of the solution studied in the works [25, 26], which consider quasi-MC, polynomial chaos and collocation schemes, and analyze convergence of the schemes in terms of overall level. However, our focus is on the analysis of the computational complexity of the multilevel interpolation algorithms, and also includes results for functionals of the solution. In particular, we prove new interpolation error bounds on functionals of the solution that are needed for the analysis of the MLSC methods. Finally, we

mention [13], which applies a multilevel technique to a best-N term approximation scheme in stochastic space coupled with a finite element discretization in the spatial domain. The convergence rates given in that work, in terms of total degrees of freedom, may be regarded as benchmarks, but are not realizable in practice.

In the independent and simultaneous work [35], results that closely resemble those of this paper are presented; the setting for both papers is the same, namely elliptic PDEs with random coefficients. However, the errors analyzed in the two papers are significantly different. In [35], the the analysis is focused on the quadrature error committed by approximating the expected value of a functional of the PDE solution by the expected value of the functional applied to an SC approximation of the PDE solution. On the other hand, in this work, we consider a generalized SC approximation to the PDE as well as a functional of the PDE solution, i.e., the order of applying the functional and the SC approximation are reversed. Moreover, the approach of this paper enables a more straightforward and somewhat sharper analysis, leading to optimal convergence rates of the error. Finally, the analysis in this paper is also valid for a wider class of isotropic and anisotropic SC methods, beyond the standard Clenshaw-Curtis-based sparse grid methods treated in [35].

The outline of the paper is as follows. In Section 2, we introduce the mathematical problem, the main notation used throughout, the assumptions on the parametrization of the random inputs that are used to transform the original stochastic problem into a deterministic parametric version, and necessary assumptions about the regularity of the solution of the PDE. A description of the spatial and stochastic approximations as well as the formulation of the MLSC method follows in Section 3. In Section 4, we provide a general convergence and complexity analysis for the MLSC method. As an example of a specific single level SC approach satisfying our interpolation assumptions, we describe, in Section 5, a generalized sparse grid stochastic collocation approach based on global Lagrange interpolation. In Section 6, we provide numerical results that illustrate the theoretical results and complexity estimates and also explore issues related to the implementation of the MLSC method.

2. Problem Setting.

Consider the problem of approximating the solution of an elliptic partial differential equation (PDE) with random input data. To this end, let $D \subset \mathbb{R}^d$, $d = 1, 2, 3$, denote a bounded, Lipschitz domain with boundary denoted by ∂D and let $(\Omega, \mathscr{F}, \mathbb{P})$ denote a complete probability space. Here, Ω denotes the set of outcomes, $\mathscr{F} \subset 2^\Omega$ the σ-algebra of events, and $\mathbb{P} : \mathscr{F} \to [0, 1]$ a probability measure. Given a random field $a(\omega, \mathbf{x}) : \Omega \times \overline{D} \to \mathbb{R}$, the model problem we consider is stated as follows: find $u(\omega, \mathbf{x}) : \Omega \times \overline{D} \to \mathbb{R}$ such that almost surely

$$\begin{cases} -\nabla(a(\omega, \mathbf{x}) \cdot \nabla u(\omega, \mathbf{x})) &= f(\mathbf{x}) \quad \text{in } D \\ u(\omega, \mathbf{x}) &= 0 \quad \text{on } \partial D. \end{cases} \tag{2.1}$$

We make the following assumptions on a:

A1. (*Finite-dimensional noise*) The random field a is determined by a finite number N of random variables, denoted by the random vector $\boldsymbol{y}(\omega) := [y_1(\omega), \dots y_N(\omega)] : \Omega \to \mathbb{R}^N$.

A2. (*Boundedness*) The image $\Gamma_n := y_n(\Omega)$ of y_n is bounded for all $n \in \{1, \dots, N\}$ and, with $\Gamma = \prod_{n=1}^N \Gamma_n$, the random variables $\boldsymbol{y}$ have a joint probability density function $\rho(\boldsymbol{y}) = \prod_{n=1}^N \widetilde{\rho}(y_n) \in L^\infty(\Gamma)$, where $\widetilde{\rho}(\cdot) : [-1, 1] \to \mathbb{R}$ denotes the one-dimensional

PDF corresponding to the probability space of the random fields. Without loss of generality, we assume that $\Gamma = [-1,1]^N$.

Remark 2.1. *Another setting having a finite number of random variables occurs when the coefficient a depends on a finite number of independent scalar random physical parameters, e.g., diffusivities, reaction rates, porosities, elastic moduli, etc. In this case, each of the N parameters would have its own PDF $\rho_n(y_n)$, $n = 1, \ldots, N$, so that the joint PDF is now given by $\rho(\boldsymbol{y}) = \prod_{n=1}^{N} \rho_n(y_n)$. The algorithms discussed in this work all apply equally well to this setting.*

Under assumptions **A1** and **A2**, the solution u to (2.1) depends measurably on a, and therefore it follows from the Doob-Dynkin Lemma that u can also be characterized in terms of the random vector $\boldsymbol{y}(\omega)$. The solution $u(\omega, \mathbf{x})$ thus has a deterministic, parametric equivalent $u(\boldsymbol{y}, \mathbf{x})$, with the probability space $(\Gamma, \mathcal{B}, \rho(\boldsymbol{y})\mathrm{d}\boldsymbol{y})$ taking the place of $(\Omega, \mathscr{F}, \mathbb{P})$; see, e.g., [1]. Here, $\mathcal{B}$ denotes the Borel σ-algebra generated by the open subsets of Γ. In what follows, we will therefore denote the solution by $u(\boldsymbol{y}, \mathbf{x})$ for $\boldsymbol{y} \in \Gamma$ and $\mathbf{x} \in D$. Then we also assume:

A3. (*Existence and uniqueness*) The coefficient $a(\omega, \mathbf{x})$ is uniformly bounded and coercive, i.e., there exists $a_{min} > 0$ and $a_{max} < \infty$ such that

$$\mathrm{Prob}\left[\omega \in \Omega : \ a_{min} \leq a(\boldsymbol{y}(\omega), \mathbf{x}) \leq a_{max} \ \ \forall\, \mathbf{x} \in \overline{D}\right] = 1$$

and $f \in H^{-1}(D)$ so that the problem (2.1) admits a unique solution $u \in L^2_\rho(\Gamma; H^1_0(D))$ with realizations in $H^1_0(D)$, i.e., $u(\boldsymbol{y}(\omega), \cdot) \in H^1_0(D)$ almost surely.

Here, given a Banach space $X(D)$ of functions on D, the weighted Bochner spaces $L^q_\rho(\Gamma; X(D))$ for $1 \leq q < \infty$ are defined by

$$L^q_\rho(\Gamma; X(D)) = \left\{ v : \Gamma \to X(D) \mid v \text{ is strongly meas. and } \int_\Gamma \|v(\boldsymbol{y}, \cdot)\|^q_{X(D)} \rho(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} < \infty \right\}$$

with corresponding norm $\|\cdot\|_{L^q_\rho(\Gamma; X(D))}$ given by

$$\|v\|^q_{L^q_\rho(\Gamma; X(D))} = \int_\Gamma \|v(\boldsymbol{y}, \cdot)\|^q_{X(D)} \rho(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}.$$

Assumption **A1** is naturally satisfied by random fields that only depend on a finite set of parameters, e.g.,

$$a(\omega, \mathbf{x}) = a(\boldsymbol{y}(\omega), \mathbf{x}) = a_0(\mathbf{x}) + \sum_{n=1}^{N} y_n(\omega) a_n(\mathbf{x}), \quad \{a_n\}_{n=0}^{N} \subset L^2(D),$$

where $\boldsymbol{y}(\omega)$ is a vector of independent random variables. If this is not the case, approximations of a that satisfy assumption **A1** can be obtained by appropriately truncating a spectral expansion such as the Karhunen-Loève expansion [13,19]. This introduces an additional error; see [31] for a discussion of the effect of this error on the convergence of stochastic collocation methods and [7,17] for bounds on the truncation error. As an alternative to truncating infinite expansions, one can also consider using dimension-adaptive sparse grids as interpolation operators. For more details on this type of approximation, we refer the reader to [9,18].

Assumption **A2** can be weakened to include the case of unbounded random variables such as Gaussian variables. See [1] for an analysis of the interpolation error and note that, with only minor modifications, the multilevel stochastic collocation method introduced in this paper also applies to unbounded random variables. Furthermore, assumption **A3** can be weakened to include coefficients a that are not uniformly coercive; see [8, 34].

Finally, we remark that the multilevel stochastic collocation method proposed in this paper is not specific to the model problem (2.1); it can be applied also to higher-order PDEs and other types of boundary conditions.

3. Hierarchical multilevel stochastic collocation methods.

We begin by recalling that standard stochastic collocation (SC) methods generally build an approximation of the solution u by evaluating a spatial approximation $u_h(\boldsymbol{y}, \cdot) \in V_h$ at a given set of points $\{\boldsymbol{y}_m\}_{m=1}^M$ in Γ, where $V_h \subset H_0^1(D)$ is a finite-dimensional subspace. In other words, we compute $\{u_h(\boldsymbol{y}_m, \cdot)\}_{m=1}^M$. Then, given a basis $\{\phi_m(\boldsymbol{y})\}_{m=1}^M$ for the space $\mathcal{P}_M = \operatorname{span}\{\phi_m(\boldsymbol{y})\}_{m=1}^M \subset L_\rho^2(\Gamma)$, we use those samples to construct the fully discrete approximation given by the interpolant

$$u_{M,h}^{(\mathrm{SL})}(\boldsymbol{y}, \mathbf{x}) = \mathcal{I}_M[u_h](\boldsymbol{y}, \mathbf{x}) = \sum_{m=1}^{M} c_m(\mathbf{x})\, \phi_m(\boldsymbol{y}), \tag{3.1}$$

where the coefficients $c_m(\mathbf{x})$ are fully determined by the semi-discrete solutions at the collocation points, $u_h(\boldsymbol{y}_m, \mathbf{x})$ for $m = 1, \ldots, M$. In (3.1), we label the standard SC approximation by 'SL' to indicate that that approximation is constructed using a single set of points $\{\boldsymbol{y}_m\}_{m=1}^M$ in stochastic space, in contrast to the multilevel approximations considered below that use a hierarchy of point sets; thus, we refer to (3.1) as a *single level* approximation. A wide range of choices for the interpolation points $\{\boldsymbol{y}_m\}_{m=1}^M$ and basis functions $\{\phi_m(\boldsymbol{y})\}_{m=1}^M$ are possible. A particular example of the approximation (3.1), namely global Lagrange interpolation on generalized sparse grids, is given in Section 5.

Convergence of the SC approximation (3.1) is often assessed in the natural $L_\rho^2(\Gamma; H_0^1(D))$-norm, and the goal is to determine a bound on the error $\|u - \mathcal{I}_M[u_h]\|_{L_\rho^2(\Gamma; H_0^1(D))}$. To obtain a good approximation with SC methods, it is necessary in general to use accurate spatial approximations u_h and a large number M of collocation points. To determine the coefficients $c_m(\mathbf{x})$ of the interpolant (3.1), the method requires the computation of $u_h(\boldsymbol{y}_m, \cdot)$ for $m = 1, \ldots, M$ so that, in practice, the cost can grow quickly with increasing M. Therefore, to reduce the overall cost, we consider a multilevel version of SC methods that combines different levels of fidelity of both the spatial and parameter approximations.

3.1. Spatial approximation.

For spatial approximation, we use a hierarchical family of finite element discretizations [5, 10]. As discussed in [26], the formulation of the multilevel method does not depend on the specific spatial discretization scheme used and the results readily hold for other choices. For $k \in \mathbb{N}_0$, define a hierarchy of nested finite element spaces

$$V_{h_0} \subset V_{h_1} \subset \cdots \subset V_{h_k} \subset \cdots \subset H_0^1(D),$$

where each V_{h_k} consists of continuous, piecewise polynomial functions on a shape regular triangulation τ_{h_k} of D having maximum mesh spacing parameter h_k. Note that k merely serves to index the given spaces; the approximation properties of the space V_{h_k} is governed

by h_k. For simplicity, we assume that the triangulations $\{\tau_{h_k}\}_{k\in\mathbb{N}_0}$ are generated by iterative uniform subdivisions of the initial triangulation τ_0; this implies that $h_k = \eta^{-k}h_0$ for some $\eta \in \mathbb{N}$, $\eta > 1$ and that indeed the corresponding finite element spaces are nested.

Remark 3.1. *For simplicity, we have assumed that the finite element family of spaces is nested, and in fact, are constructed by a series of uniform subdivisions of a parent mesh with mesh size h_0. Neither of these assumptions are necessary for our algorithms or conclusions to hold, provided $\eta_1 \le h_k/h_{k+1} \le \eta_2$ for some $0 < \eta_1 < \eta_2 < \infty$ and all $k \in \mathbb{N}_0$; in such cases, the finite element spaces are not necessarily nested.*

We also let $u_{h_k}(\boldsymbol{y}, \cdot)$ denote the Galerkin projection of $u(\boldsymbol{y}, \cdot)$ onto V_{h_k}, i.e., $u_{h_k} \in V_{h_k}$ denotes the finite element approximation. Note that $u_{h_k}(\boldsymbol{y}, \cdot)$ is still a function on the stochastic parameter space Γ. We assume the following approximation property of the finite element spaces $\{V_{h_k}\}_{k\in\mathbb{N}_0}$:

A4. There exist positive constants α and C_s, independent of h_k, such that for all $k \in \mathbb{N}_0$,

$$\|u - u_{h_k}\|_{L^2_\rho(\Gamma;H^1_0(D))} \le C_s\, h_k^\alpha.$$

In general, the rate α depends on the (spatial) regularity of u, which in turn depends on the regularity of a and f as well as on the geometry of the domain D. For example, if a, f, and D are sufficiently regular so that $u \in L^2_\rho(\Gamma; H^2(D))$, assumption **A4** holds with $\alpha = 1$ and C_s dependent only on a and $\|u\|_{L^2_\rho(\Gamma;H^2(D))}$. For additional examples and detailed analyses of finite element errors, see [34].

3.2. Stochastic interpolation. For stochastic approximation, we use interpolation over Γ, where we assume $u \in C^0(\Gamma; H^1_0(D))$. The specific choice of interpolation scheme is not crucial at this juncture. We begin by letting $\{\mathcal{I}_{M_k}\}_{k=0}^\infty$ denote a sequence of interpolation operators $\mathcal{I}_{M_k} : C^0(\Gamma) \to L^2_\rho(\Gamma)$ using M_k points. We assume the following:

A5. There exist positive constants C_I, C_ζ, and β, and a Banach space $\Lambda(\Gamma; H^1_0(D)) \subset L^2_\rho(\Gamma; H^1_0(D))$ containing the finite element approximations $\{u_{h_k}\}_{k\in\mathbb{N}_0}$ such that for all $v \in \Lambda(\Gamma; H^1_0(D))$ and all $k \in \mathbb{N}_0$

$$\|v - \mathcal{I}_{M_k} v\|_{L^2_\rho(\Gamma;H^1_0(D))} \le C_I\, \sigma(M_k)\, \zeta(v),$$

for some decreasing sequence $\{\sigma_k\}_{k\in\mathbb{N}_0}$, with $\sigma_k = \sigma(M_k)$, and operator $\zeta : \Lambda(\Gamma; H^1_0(D)) \to \mathbb{R}$ that admits the estimates

$$\zeta(u_{h_k}) \le C_\zeta\, h_0^\beta \quad \text{and} \quad \zeta(u_{h_{k+1}} - u_{h_k}) \le C_\zeta\, h_{k+1}^\beta.$$

Remark 3.2. *As in the previous section, k is merely an index; we use the same index for the hierarchies of spatial and stochastic approximations because, in the multilevel SC method we introduce below, these two hierarchies are closely connected.*

Remark 3.3. *σ_k determines the approximation properties of the interpolant. Moreover, we allow non-unique interpolation operators in the sequence, i.e., it is possible that, for any $k = 0, \ldots, \infty$, $M_{k+1} = M_k$ and therefore $\mathcal{I}_{M_{k+1}} = \mathcal{I}_{M_k}$ and $\sigma_{k+1} = \sigma_k$. Thus, although the spatial approximation improves with increasing k, i.e., $h_{k+1} < h_k$, we allow for the parameter space approximation for the index $k+1$ remaining the same as that for k.*

In Section 5, assumption **A5** is shown to hold for global Lagrange interpolation using generalized sparse grids, with $\sigma_k = M_k^{-\mu}$, with Λ a space of $H_0^1(D)$-valued functions over Γ admitting a complex extension in a suitable region $\mathcal{E} \in \mathbb{C}^N$, and defining $\zeta(v) = \sup_{\boldsymbol{z}\in\mathcal{E}} \|v(\boldsymbol{z})\|_{H_0^1(D)}$. The bounds on the function ζ in assumption **A5** are shown to be the key to balancing spatial and stochastic discretizations through the multilevel formulation. Crucially, we make use of the fact that the interpolation error is proportional to the size of the function being interpolated, measured in an appropriate norm. In the case of the model problem (2.1), this norm is usually related to the (spatial) $H_0^1(D)$-norm. The bounds in assumption **A5** then arise from the fact that for any $k \in N_0$, $\|u_{h_k}\|_{H_0^1(D)}$ is bounded by a constant, independent of k, whereas $\|u_{h_k} - u_{h_{k-1}}\|_{H_0^1(D)}$ decays with h_k^β for some $\beta > 0$. We usually have $\beta = \alpha$, where α is as in assumption **A4**. Note that we have chosen to scale the bound on $\zeta(u_{h_k})$ by h_0^β to simplify calculations. Because h_0 is a constant, this does not affect the nature of the assumption.

3.3. Formulation of the multilevel method. As in the previous sections, denote by $\{u_{h_k}\}_{k\in\mathbb{N}_0}$ and $\{\mathcal{I}_{M_k}\}_{k\in\mathbb{N}_0}$ sequences of spatial approximations and interpolation operators in parameter space, respectively. Then, for any $K \in \mathbb{N}$, the formulation of the multilevel method begins with the simple telescoping identity

$$u_{h_K} = \sum_{k=0}^{K} (u_{h_k} - u_{h_{k-1}}), \tag{3.2}$$

where, for simplicity, we set $u_{h_{-1}} := 0$.

It follows from assumption **A5** that as $k \to \infty$, less accurate interpolation operators are needed in order to estimate $u_{h_k} - u_{h_{k-1}}$ to achieve a required accuracy. We therefore define our multilevel interpolation approximation as

$$u_K^{(\mathrm{ML})} := \sum_{k=0}^{K} \mathcal{I}_{M_{K-k}}[u_{h_k} - u_{h_{k-1}}] = \sum_{k=0}^{K} \left(u^{(\mathrm{SL})}_{M_{K-k},h_k} - u^{(\mathrm{SL})}_{M_{K-k},h_{k-1}} \right). \tag{3.3}$$

Rather than simply interpolating u_{h_K}, this approximation uses different levels of interpolation on each difference $u_{h_k} - u_{h_{k-1}}$ of finite element approximations. To preserve convergence, the estimator uses the most accurate interpolation operator $\mathcal{I}_{M_K}$ on the coarsest spatial approximation u_{h_0} and the least accurate interpolation operator $\mathcal{I}_{M_0}$ on the finest spatial approximation $u_{h_K} - u_{h_{K-1}}$. Note that in (3.3) a single index k is used to select appropriate spatial and stochastic approximations and thus these approximations are indeed closely related.

4. Analysis of the multilevel approximation. This section is devoted to proving the convergence of the multilevel approximation defined in Section 3.3 and analyzing its computational complexity. We first prove, in Section 4.1, a general error bound, whereas in Sections 4.2 and 4.3 we prove a bound on the computational complexity in the particular case of an algebraic decay of the interpolation errors.

4.1. Convergence analysis. We consider the convergence of the multilevel approximation $u_K^{(\mathrm{ML})}$ to the true solution u in the natural norm $\|\cdot\|_{L^2_\rho(\Gamma;H_0^1(D))}$.

First, we use the triangle inequality to split the error into the sum of a spatial discretization error and a stochastic interpolation error, i.e.,

$$\|u - u_K^{(\mathrm{ML})}\|_{L^2_\rho(\Gamma;H^1_0(D))} \leq \underbrace{\|u - u_{h_K}\|_{L^2_\rho(\Gamma;H^1_0(D))}}_{(I)} + \underbrace{\|u_{h_K} - u_K^{(\mathrm{ML})}\|_{L^2_\rho(\Gamma;H^1_0(D))}}_{(II)}. \tag{4.1}$$

The aim is to prove that with the interpolation operators $\{\mathcal{I}_{M_k}\}_{k=0}^K$ chosen appropriately, the stochastic interpolation error (II) of the multilevel approximation converges at the same rate as the spatial discretization error (I), hence resulting in a convergence result for the total error.

For the spatial discretization error (I), it follows immediately from assumption **A4** that

$$(I) \leq C_s h_K^\alpha.$$

From (3.2) and assumption **A5**, we estimate the stochastic interpolation error using the triangle inequality:

$$\begin{aligned}
(II) &= \Big\| \sum_{k=0}^{K} (u_{h_k} - u_{h_{k-1}}) - \mathcal{I}_{M_{K-k}}(u_{h_k} - u_{h_{k-1}}) \Big\|_{L^2_\rho(\Gamma;H^1_0(D))} \\
&\leq \sum_{k=0}^{K} \big\| (u_{h_k} - u_{h_{k-1}}) - \mathcal{I}_{M_{K-k}}(u_{h_k} - u_{h_{k-1}}) \big\|_{L^2_\rho(\Gamma;H^1_0(D))} \\
&\leq \sum_{k=0}^{K} C_I\, C_\zeta\, \sigma_{K-k}\, h_k^\beta.
\end{aligned}$$

To obtain an error of the same size as (I), we choose interpolation operators such that

$$\sigma_{K-k} \leq C_s \left((K+1)\, C_I\, C_\zeta \right)^{-1} h_K^\alpha\, h_k^{-\beta}. \tag{4.2}$$

Continuing from above,

$$(II) \leq \sum_{k=0}^{K} C_s \left((K+1)\, C_I\, C_\zeta \right)^{-1} h_K^\alpha\, h_k^{-\beta} C_I\, C_\zeta\, h_k^\beta = C_s h_K^\alpha,$$

as required. It follows that with σ_k as in (4.2)

$$\|u - u_K^{(\mathrm{ML})}\|_{L^2_\rho(\Gamma;H^1_0(D))} \leq 2\, C_s\, h_K^\alpha.$$

4.2. Cost analysis. We now proceed to analyze the computational cost of the MLSC method. We consider the *ε-cost of the estimator*, denoted here by $C_\varepsilon^{\mathrm{ML}}$, which is the computational cost required to achieve a desired accuracy ε. In order to quantify this cost, we use the convergence rates of the spatial discretization error and, for the stochastic interpolation error, the rates given by assumptions **A4** and **A5**. In particular, we will assume that **A5** holds with $\sigma_k = M_k^{-\mu}$ for some $\mu > 0$.

Remark 4.1. *The choice $\sigma_k = M_k^{-\mu}$ best reflects approximations based on SC methods that employ sparse grids. In particular, as mentioned in Section 3.2, algebraic decay holds for the generalized sparse grid interpolation operators considered in Section 5; see Theorem 5.5. For other possible choices in the context of quadrature, see [26].*

In general, the MLSC method involves solving, for each k, the deterministic PDE for each of the M_k sample points from Γ; in fact, according to (3.3), two solves are needed, one for each of two spatial grid levels. Thus, we also require a bound on the cost, which we denote by C_k, of computing $u_{h_k} - u_{h_{k-1}}$ at a sample point. We assume:

A6. There exist positive constants γ and C_c, independent of h_k, such that $C_k \leq C_c\, h_k^{-\gamma}$ for all $k \in \mathbb{N}_0$.

If an optimal linear solver is used to solve the finite element equations for u_{h_k}, this assumption holds with $\gamma \approx d$ (see, e.g., [5]), where d is the spatial dimension. Note that the constant C_c will in general depend on the refinement ratio η described in Section 3.1.

We quantify the total computational cost of the MLSC approximation (3.3) using the metric

$$C^{(\mathrm{ML})} = \sum_{k=0}^{K} M_{K-k}\, C_k. \tag{4.3}$$

We now have the following result for the ε-cost of the MLSC method required to achieve an accuracy $\|u - u_K^{(\mathrm{ML})}\|_{L^2_\rho(\Gamma;H^1_0(D))} \leq \varepsilon$. In the analysis, we define the relations $a \lesssim b$ and $a \eqsim b$ to indicate that $a \leq Cb$ (resp. $a = Cb$) for some constant C independent the mesh width h, the number of interpolation points M and the accuracy ε.

Theorem 4.2. *Suppose assumptions* **A4**–**A6** *hold with $\sigma_k = M_k^{-\mu}$, and assume that $\alpha \geq \min(\beta, \mu\gamma)$. Then, for any $\varepsilon < \exp[-1]$, there exists an integer K, and a sequence $\{M_k\}_{k=0}^K$, such that*

$$\|u - u_K^{(\mathrm{ML})}\|_{L^2_\rho(\Gamma;H^1_0(D))} \leq \varepsilon$$

and

$$C_\varepsilon^{(\mathrm{ML})} \lesssim \begin{cases} \varepsilon^{-\frac{1}{\mu}}, & \text{if } \beta > \mu\gamma \\ \varepsilon^{-\frac{1}{\mu}} |\log \varepsilon|^{1+\frac{1}{\mu}} & \text{if } \beta = \mu\gamma \\ \varepsilon^{-\frac{1}{\mu} - \frac{\gamma\mu - \beta}{\alpha\mu}} & \text{if } \beta < \mu\gamma. \end{cases} \tag{4.4}$$

Proof. As in (4.1), we consider separately the two error contributions (I) and (II). To achieve the desired accuracy, it is sufficient to bound both error contributions by $\frac{\varepsilon}{2}$. Without loss of generality, for the remainder of this proof we assume $h_0 = 1$. If this is not the case, we simply need to rescale the constants C_s, C_ζ, and C_c.

First, we choose K large enough so that $(I) \leq \frac{\varepsilon}{2}$. By assumption **A4**, it is sufficient to require $C_s h_K^\alpha \leq \frac{\varepsilon}{2}$. Because the hierarchy of meshes $\{h_k\}_{k\in\mathbb{N}_0}$ is obtained by uniform refinement, $h_k = \eta^{-k} h_0 = \eta^{-k}$, and we have

$$h_K \leq \left(\frac{\varepsilon}{2C_s}\right)^{1/\alpha} \quad \text{if} \quad K = \left\lceil \frac{1}{\alpha} \log_\eta \left(\frac{2C_s}{\varepsilon}\right) \right\rceil. \tag{4.5}$$

This fixes the total number of levels K.

In order to obtain the multilevel estimator with the smallest computational cost, we now determine the $\{M_k\}_{k=0}^K$ so that the computational cost (4.3) is minimized, subject to the requirement $(II) \leq \frac{\varepsilon}{2}$. Treating the M_k as continuous variables, we use the Lagrange multiplier method. To begin, we form the Lagrange function, using assumptions **A4**-**A6**.

$$\mathcal{L}(M_0, \ldots, M_K, \lambda) = \sum_{k=0}^{K} C_c \, M_{K-k} \, \eta^{k\gamma} + \lambda \left(\sum_{k=0}^{K} C_I \, C_\zeta \, M_{K-k}^{-\mu} \, \eta^{-k\beta} - \varepsilon/2 \right).$$

Here we have replaced the condition $(II) \leq \varepsilon/2$ with the sufficient condition $\sum_{k=0}^{K} C_I \, C_\zeta \, M_{K-k}^{-\mu} \, \eta^{-k\beta} \leq \varepsilon/2$. To find a relative extremum, we require $\nabla \mathcal{L} = 0$, leading to the $K+2$ conditions

$$\frac{\partial \mathcal{L}}{\partial M_{K-k}} = \eta^{k\gamma} - \lambda C_I \, C_\zeta \mu M_{K-k}^{-(\mu+1)} \eta^{-k\beta} = 0, \quad k = 0, \ldots, K, \tag{4.6}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{k=0}^{K} C_I \, C_\zeta \, M_{K-k}^{-\mu} \, \eta^{-k\beta} - \varepsilon/2 = 0. \tag{4.7}$$

Solving the first $K+1$ equations (4.6) for M_{K-k} yields

$$M_{K-k} = (C_I \, C_\zeta \mu \lambda)^{1/(\mu+1)} \eta^{\frac{-k(\beta+\gamma)}{\mu+1}}, \quad k = 0, \ldots, K. \tag{4.8}$$

Now, substitute (4.8) into (4.7), and solve for λ to obtain

$$\lambda = (2^{\mu+1} C_I C_\zeta)^{1/\mu} \mu^{-1} \varepsilon^{-(\mu+1)/\mu} \mathcal{S}(\eta, K)^{(\mu+1)/\mu},$$

where

$$\mathcal{S}(\eta, K) = \sum_{k=0}^{K} \eta^{-k(\frac{\beta-\gamma\mu}{\mu+1})}.$$

Inserting this into (4.8) results in the optimal choice

$$M_{K-k} = \left(2 \, C_I \, C_\zeta \, \mathcal{S}(\eta, K)\right)^{1/\mu} \varepsilon^{-1/\mu} \, \eta^{-\frac{k(\beta+\gamma)}{\mu+1}}. \tag{4.9}$$

Because M_{K-k} given by (4.9) is, in general, not an integer, we choose

$$M_{K-k} = \left\lceil (2 \, C_I \, C_\zeta \, \mathcal{S}(\eta, K))^{1/\mu} \, \varepsilon^{-1/\mu} \, \eta^{-\frac{k(\beta+\gamma)}{\mu+1}} \right\rceil. \tag{4.10}$$

Note that this choice determines the sequence $\{M_k\}_{k=0}^K$ and consequently $\{\mathcal{I}_{M_k}\}_{k=0}^K$. Also note that, in practice, this choice may not be possible for all interpolation schemes; see Remark 4.4.

With the number of samples M_{K-k} fixed, we now examine the complexity of the multilevel approximation. Since $\lceil x \rceil < x + 1$, for any $x \in \mathbb{R}$, we have

$$\begin{aligned} C_\varepsilon^{(\mathrm{ML})} &= \sum_{k=0}^{K} M_{K-k} C_k \lesssim \sum_{k=0}^{K} M_{K-k}\, \eta^{k\gamma} \\ &\lesssim \sum_{k=0}^{K} \Big(\frac{\varepsilon}{\mathcal{S}(\eta, K)}\Big)^{-\frac{1}{\mu}} \eta^{-k\frac{\beta+\gamma}{\mu+1}} \eta^{k\gamma} + \sum_{k=0}^{K} \eta^{k\gamma} \\ &\eqsim \varepsilon^{-\frac{1}{\mu}} \mathcal{S}(\eta, K)^{\frac{1}{\mu}} \sum_{k=0}^{K} \eta^{-k\frac{\beta+\gamma-\gamma(\mu+1)}{\mu+1}} + \sum_{k=0}^{K} \eta^{k\gamma} \\ &\eqsim \varepsilon^{-\frac{1}{\mu}} \mathcal{S}(\eta, K)^{\frac{1}{\mu}} \sum_{k=0}^{K} \eta^{-k\frac{\beta-\gamma\mu}{\mu+1}} + \sum_{k=0}^{K} \eta^{k\gamma} \\ &\eqsim \varepsilon^{-\frac{1}{\mu}} \mathcal{S}(\eta, K)^{1+\frac{1}{\mu}} + \sum_{k=0}^{K} \eta^{k\gamma}. \end{aligned} \tag{4.11}$$

To bound the cost in terms of ε, first note that because $K < \frac{1}{\alpha} \log_\eta(2C_s/\varepsilon) + 1$ by (4.5), we have

$$\sum_{k=0}^{K} \eta^{k\gamma} \leq \frac{\eta^{\gamma K}}{1 - \eta^{-\gamma}} \leq \frac{\eta^{\gamma} (2C_s)^{\gamma/\alpha}}{1 - \eta^{-\gamma}} \varepsilon^{-\gamma/\alpha}. \tag{4.12}$$

Next, we need to consider different values of β and μ. When $\beta > \gamma\mu$, $\mathcal{S}(\eta, K)$ is a geometric sum that converges to a limit independent of K. Because $\alpha \geq \gamma\mu$ implies that $\varepsilon^{-\gamma/\alpha} \leq \varepsilon^{-\frac{1}{\mu}}$ for $\varepsilon < \exp[-1] < 1$, we have $C_\varepsilon^{(\mathrm{ML})} \lesssim \varepsilon^{-\frac{1}{\mu}}$ in this case.

When $\beta = \gamma\mu$, we find that $\mathcal{S}(\eta, K) = K+1$, and so, using (4.5), $\alpha \geq \mu\gamma$, and $\varepsilon < \exp[-1]$,

$$C_\varepsilon^{(\mathrm{ML})} \lesssim \varepsilon^{-\frac{1}{\mu}} (K+1)^{1+\frac{1}{\mu}} + \varepsilon^{-\frac{\gamma}{\alpha}} \eqsim \varepsilon^{-\frac{1}{\mu}} |\log \varepsilon|^{1+\frac{1}{\mu}}.$$

For the final case of $\beta < \gamma\mu$, we reverse the index in the sum $\mathcal{S}(\eta, K)$ to obtain a geometric sequence

$$\mathcal{S}(\eta, K) = \sum_{k=0}^{K} \eta^{(k-K)\frac{\beta-\gamma\mu}{\mu+1}} = \eta^{-K\frac{\beta-\gamma\mu}{\mu+1}} \sum_{k=0}^{K} \eta^{-k(\frac{\gamma\mu-\beta}{\mu+1})} \lesssim \varepsilon^{\frac{\beta-\gamma\mu}{\alpha(\mu+1)}}.$$

Because $\alpha \geq \beta$, this gives

$$C_\varepsilon^{(\mathrm{ML})} \lesssim \varepsilon^{-\frac{1}{\mu}} \varepsilon^{\frac{\beta-\gamma\mu}{\alpha(\mu+1)}(1+\frac{1}{\mu})} + \varepsilon^{-\frac{\gamma}{\alpha}} \eqsim \varepsilon^{-\frac{1}{\mu} - \frac{\gamma\mu-\beta}{\alpha\mu}}.$$

This completes the proof. ■

Remark 4.3. **Assumptions.** *The requirement $\alpha \geq \min(\beta, \mu\gamma)$ is not restrictive and may be easily verified in practice, since most often $\alpha = \beta$. Indeed, this assumption is satisfied for each of the examples in Section 6.*

Remark 4.4. **Error and quadrature level.** *In this section, we characterized the convergence of the interpolation errors in terms of the number of interpolation points M_k. Yet when*

computing quadratures based on sparse grid techniques (see Section 5), an arbitrary number of points will not in general have an associated sparse grid. Thus, choosing an interpolant using the optimal number of points according to (4.10) may not be possible in practice. However, in light of estimates such as [31, Lemma 3.9], it is not unreasonable to make the assumption that given any number of points M, there exists an interpolant using $\widetilde{M}$ points, with

$$M \leq \widetilde{M} \leq C M^{\delta} \tag{4.13}$$

for some $\delta \geq 1$. We can think of δ as measuring the inefficiency of our sparse grids in representing higher-dimensional polynomial spaces. Using (4.13), one can proceed as in Theorem 4.2 to derive a bound on the ε-cost of the resulting multilevel approximation.

Another possibility would be to solve a discrete, constrained minimization problem to find optimal interpolation levels, relying on convergence results for the interpolation error in terms of the interpolation level rather than number of points; see [32, Theorem 3.4]. However, our cost metric relies on precise knowledge of the number of points, making theoretical comparison difficult.

Remark 4.5. **Cancellations and computational cost.** *The cost estimate (4.3) takes into consideration the cost of all the terms in the multilevel estimator (3.3). However, when the same interpolation operator is used on two consecutive levels, terms in the multilevel approximation cancel and need in fact not be computed. For example, if $\mathcal{I}_{M_{K-k}} = \mathcal{I}_{M_{K-k-1}}$, then*

$$\mathcal{I}_{M_{K-k}}(u_{h_k} - u_{h_{k-1}}) + \mathcal{I}_{M_{K-k-1}}(u_{h_{k+1}} - u_{h_k}) = \mathcal{I}_{M_{K-k}}(u_{h_{k+1}} - u_{h_{k-1}})$$

so that the computation of the interpolants of u_{h_k} is not necessary. Especially in the context of sparse grid interpolation, in practice we choose the same interpolation grid for several consecutive levels, leading to a significant reduction in the actual computational cost compared to that estimated in Theorem 4.2. The effect of these cancellations is clearly visible in some of the numerical experiments of Section 6, see the discussion of Figures 6.2 and 6.5 for more details.

4.2.1. Comparison to single level collocation methods. Under the same assumptions as in Theorem 4.2, for any $M_{sl} \in \mathbb{N}_0$ and h_{sl}, the error in the standard single-level SC approximation (3.1) can be bounded by

$$\|u - u^{(\mathrm{SL})}_{M_{sl},h_{sl}}\|_{L^2_\rho(\Gamma;H^1_0(D))} \leq C_s\, h_{sl}^{\alpha} + C_I\, \zeta(u_h)\, M_{sl}^{-\mu}.$$

To make both contributions equal to $\varepsilon/2$, it suffices to choose $h_{sl} \eqsim \varepsilon^{1/\alpha}$ and $M_{sl} \eqsim \varepsilon^{-1/\mu}$. This choice determines M_{sl} and hence $\mathcal{I}_{M_{sl}}$. The computational cost to achieve a total error of ε is then bounded by

$$C^{(\mathrm{SL})}_{\varepsilon} \eqsim h_{sl}^{-\gamma} M_{sl} \eqsim \varepsilon^{-\frac{1}{\mu}-\frac{\gamma}{\alpha}}.$$

A comparison with the bounds on computational complexity proved in Theorem 4.2 shows clearly the superiority of the multilevel method.

In the case $\beta > \gamma\mu$, the convergence rate of the finite element correction errors is comparatively larger than the convergence rate of the interpolant when multiplied by the cost factor γ. From (4.11), this indicates that the cost $M_{K-k}C_k$ is largest at the coarsest level $k = 0$, and

hence most of the computational effort of the multilevel approximation is expended computing $\mathcal{I}_{M_K}(u_{h_0})$. The savings in cost compared to single level SC hence correspond to the difference in cost between obtaining samples u_{h_0} on the coarse grid h_0 and obtaining samples u_{h_K} on the fine grid $h_{sl} = h_K$ used by the single-level method. This gives a saving of $(h_{sl}/h_0)^\gamma \eqsim \varepsilon^{\gamma/\alpha}$.

The case $\beta = \mu\gamma$ corresponds to the computational effort being spread evenly across the levels, and, up to a log factor, the savings in cost are again of order $\varepsilon^{\gamma/\alpha}$.

In contrast, when $\beta < \gamma\mu$, i.e., when the interpolation error is converging quickly compared to the finite element approximations, the computational cost of computing one sample of u_{h_k} grows comparatively quickly with respect to k, and most of the computational effort of the multilevel approximation is on the finest level $k = K$. The benefits compared to single level SC hence corresponds approximately to the difference between M_K and M_{sl}. This gives a savings of $M_K/M_{sl} \eqsim (h_K^\beta)^{1/\mu} \eqsim \varepsilon^{\beta/\alpha\mu}$.

4.3. Multilevel approximation of functionals. In applications, it is often of interest to bound the error in the expected value of a functional ψ of the solution u, where $\psi : H_0^1(D) \to \mathbb{R}$. Similar to (3.1), the SC approximation of $\psi(u)$ is given by

$$\psi_{M,h}^{(\mathrm{SL})}[u] = \mathcal{I}_M\left[\psi(u_h)\right] \tag{4.14}$$

and, similar to (3.3), the multilevel interpolation approximation of $\psi(u)$ is given by

$$\psi_K^{(\mathrm{ML})}[u] := \sum_{k=0}^{K} \mathcal{I}_{M_{K-k}}\big(\psi(u_{h_k}) - \psi(u_{h_{k-1}})\big), \tag{4.15}$$

where, as before, we set $u_{h_{-1}} := 0$ and we also assume, without loss of generality, that $\psi(0) = 0$. Note that in the particular case of linear functionals ψ, we in fact have

$$\psi_{M,h}^{(\mathrm{SL})}[u] = \psi(u_{M,h}^{(\mathrm{SL})}) \quad \text{and} \quad \psi_K^{(\mathrm{ML})}[u] = \psi(u_K^{(\mathrm{ML})}).$$

Analogous to Theorem 4.2, we have the following result about the ε-cost for the error $\left|\mathbb{E}\big[\psi(u) - \psi_K^{(\mathrm{ML})}[u]\big]\right|$ in the expected value of the multilevel approximation of functionals.

Proposition 4.6. *Suppose there exist positive constants $\alpha, \beta, \mu, \gamma, C_s, C_I, C_\zeta, C_c$, with $\alpha \geq \min(\beta, \mu\gamma)$, and an operator $\zeta : \Lambda(\Gamma;\mathbb{R}) \to \mathbb{R}$, for a Banach space $\Lambda(\Gamma;\mathbb{R}) \subset L^2_\rho(\Gamma;\mathbb{R})$ containing the finite element approximations $\{\psi(u_{h_k})\}_{k\in\mathbb{N}_0}$, such that for all $k \in \mathbb{N}_0$ we have*

- **F1.** $|\mathbb{E}[\psi(u) - \psi(u_{h_k})]| \leq C_s\, h_k^\alpha$
- **F2.** $\left|\mathbb{E}\big[\psi(u_{h_k}) - \psi(u_{h_{k-1}}) - \mathcal{I}_{M_{K-k}}(\psi(u_{h_k}) - \psi(u_{h_{k-1}}))\big]\right| \leq C_I\, M_{K-k}^{-\mu}\, \zeta(\psi(u_{h_k}) - \psi(u_{h_{k-1}}))$
- **F3.** $\zeta(\psi(u_{h_k}) - \psi(u_{h_{k-1}})) \leq C_\zeta\, h_k^\beta$
- **F4.** $C_k \leq C_c\, h_k^{-\gamma}$.

Then, for any $\varepsilon < \exp[-1]$, there exists an integer K and a sequence $\{M_k\}_{k=0}^K$ such that

$$\left|\mathbb{E}\big[\psi(u) - \psi_K^{(\mathrm{ML})}(u)\big]\right| \leq \varepsilon,$$

with computational cost $C_\varepsilon^{(\mathrm{ML})}$ bounded as in Theorem 4.2.

The assumptions **F1**–**F4** are essentially the same as the assumptions **A4**–**A6** of Theorem 4.2, with perhaps different values for the constants C_s, C_I, C_ζ, and C_c. Certainly, bounded linear functionals have this inheritance property. In Section 5, we give some examples of nonlinear functionals that also have this property.

5. Global multi-dimensional interpolation.

In this section, we provide a specific example of a single level SC approach, given by (3.1), that will be used to construct the interpolation operators in our MLSC approach. As such, we briefly recall generalized multi-dimensional (including sparse grid) interpolation, as well as theoretical results related to the interpolation operator. For a more thorough description, see [1, 3, 31, 32].

Remark 5.1. *In this section, we again introduce a second notion of* levels. *The levels here should not be confused with the levels used previously. For the latter, 'levels' refer to members of hierarchies of spatial and stochastic approximations, both of which were indexed by k. In this section, 'levels' refer to a sequence, indexed by l, of stochastic polynomial spaces and corresponding point sets used to construct a specific sparse grid interpolant. The result of this construction, i.e., of using the levels indexed by l, is the interpolants used in the previous sections that were indexed by k.*

5.1. Construction of generalized sparse grid interpolant.

The construction of the interpolant in the N-dimensional space $\Gamma = \prod_{n=1}^{N} \Gamma_n$ is based on sequences of one-dimensional Lagrange interpolation operators $\{\mathcal{U}_n^{p(l)}\}_{l\in\mathbb{N}} : C^0(\Gamma_n) \to \mathcal{P}_{p(l)-1}(\Gamma_n)$, where $\mathcal{P}_p(\Gamma_n)$ denotes the space of polynomials of degree p on Γ_n. In particular, for each $n = 1, \ldots, N$, let $l \in \mathbb{N}_+$ denote the one-dimensional level of approximation and let $\{y_{n,j}^{(l)}\}_{j=1}^{p(l)} \subset \Gamma_n$ denote a sequence of one-dimensional interpolation points in Γ_n. Here, $p(l) : \mathbb{N}_+ \to \mathbb{N}_+$ is such that $p(1) = 1$ and $p(l) < p(l+1)$ for $l = 2, 3, \ldots$, so that $p(l)$ strictly increases with l and defines the total number of collocation points at level l. For a univariate function $v \in C^0(\Gamma_n)$, we define $\mathcal{U}_n^{p(l)}$ by

$$\mathcal{U}_n^{p(l)}[v](y_n) = \sum_{j=1}^{p(l)} v\big(y_{n,j}^{(l)}\big)\varphi_{n,j}^{(l)}(y_n) \quad \text{for } l_n = 1, 2, \ldots, \tag{5.1}$$

where $\varphi_{n,j}^{(l)} \in \mathcal{P}_{p(l)-1}(\Gamma_n)$, $j = 1, \ldots, p(l)$, are Lagrange fundamental polynomials of degree $p(l) - 1$, which are completely determined by the property $\varphi_{n,j}^{(l)}(y_{n,i}^{(l)}) = \delta_{i,j}$.

Using the convention that $\mathcal{U}_n^{p(0)} = 0$, we introduce the difference operator given by

$$\Delta_n^{p(l)} = \mathcal{U}_n^{p(l)} - \mathcal{U}_n^{p(l-1)}. \tag{5.2}$$

For the multivariate case, we let $\boldsymbol{l} = (l_1, \ldots, l_N) \in \mathbb{N}_+^N$ denote a multi-index and $L \in \mathbb{N}_+$ denote the total level of the sparse grid approximation. Now, from (5.2), the *L-th level generalized sparse-grid approximation* of $v \in C^0(\Gamma)$ is given by

$$\mathcal{A}_L^{p,g}[v] = \sum_{g(\boldsymbol{l})\le L} \bigotimes_{n=1}^{N} \Delta_n^{p(l_n)}[v], \tag{5.3}$$

where $g : \mathbb{N}_+^N \to \mathbb{N}$ is another strictly increasing function that defines the mapping between the multi-index $\boldsymbol{l}$ and the level L used to construct the sparse grid. The single level approximation

(5.3) requires the independent evaluation of v on a deterministic set of *distinct collocation points* given by

$$\mathcal{H}_L^{p,g} = \bigcup_{L-N+1\le g(\boldsymbol{l})\le L} \prod_{1\le n\le N} \left\{ y_{n,j}^{(l_n)} \right\}_{j=1}^{p(l_n)}$$

having cardinality M_L.

Remark 5.2. *For the MLSC method, the interpolation operators $\mathcal{I}_{M_k}$ introduced in Section 3.2 are chosen as $\mathcal{A}_L^{p,g}$ with $M_k = M_L$. However, we have already noted in Remark 4.4 that an arbitrary number of points will not in general have an associated sparse grid, and in practice a rounding strategy has to be applied to choose the interpolation operator on each level. For examples of rounding strategies, see the numerical examples in Section 6. Note that although in theory this rounding may change the computational complexity of the MLSC estimators, our numerical investigations confirm that the complexities proved in Theorem 4.2 are a good fit in practice.*

The particular choices of the one-dimensional growth rate $p(l)$ and the function $g(\boldsymbol{l})$ define a general multi-index set $\mathcal{J}^{p,g}(L) = \{\boldsymbol{l} : g(\boldsymbol{l}) \le L\}$ used in the construction of the sparse grid, and the corresponding underlying isotropic polynomial space of the approximation denoted $\mathcal{P}_{\mathcal{J}^{p,g}(L)}(\Gamma)$ [3,24]. Some examples of functions $p(l)$ and $g(\boldsymbol{l})$ and $\mathcal{P}_{\mathcal{J}^{p,g}(L)}(\Gamma)$ are given in Table 5.1. Extensions to anisotropic versions can be constructed by introducing a weight vector as described in Remark 5.7.

Table 5.1

The functions $p : \mathbb{N}_+ \to \mathbb{N}_+$ and $g : \mathbb{N}_+^N \to \mathbb{N}$ and the corresponding isotropic polynomial subspaces.

Polynomial Space	$p(l)$	$g(\boldsymbol{l})$
Tensor product	$p(l) = l$	$\max_{1\le n\le N}(l_n - 1)$
Total degree	$p(l) = l$	$\sum_{n=1}^N (l_n - 1)$
Hyperbolic cross	$p(l) = l$	$\prod_{n=1}^N (l_n - 1)$
Sparse Smolyak	$p(l) = 2^{l-1} + 1,\ l > 1$	$\sum_{n=1}^N (l_n - 1)$

Table 5.1 defines several polynomial spaces. A means for constructing a basis for polynomial subspaces consists of selecting a set of points and then defining basis functions based on those points, e.g., Lagrange fundamental polynomials. For sparse Smolyak polynomial spaces, the most popular choice of points are the sparse grids based on the one-dimensional Clenshaw-Curtis abscissas [11] which are the extrema of Chebyshev polynomials, including the end-point extrema. For level l, and in the particular case $\Gamma_n = [-1, 1]$ and $p(l) > 1$, the resulting points are given by

$$y_{n,j}^{(l)} = -\cos\left(\frac{\pi(j-1)}{p(l)-1}\right) \quad \text{for } j = 1, \ldots, p(l).$$

In particular, the choice $p(l)$ given in Table 5.1 for the sparse Smolyak case results in a *nested* family of one-dimensional abscissas, i.e., $\{y_{n,j}^{(l)}\}_{j=1}^{p(l)} \subset \{y_{n,j}^{(l+1)}\}_{j=1}^{p(l+1)}$, so that the sparse grids

are also nested, i.e., $\mathcal{H}_L^{p,g} \subset \mathcal{H}_{L+1}^{p,g}$. Using $g(\boldsymbol{l})$ in (5.3), given as in Table 5.1 for the Smolyak polynomial space, corresponds to the most widely used sparse-grid approximation, as first described in [33]. Other nested families of sparse grids can be constructed from, e.g., the Newton-Cotes and Gauss-Patterson one-dimensional abscissas.

Remark 5.3. *In general, the growth rate $p(l)$ can be chosen as any increasing function on $\mathbb{N}$. However, to construct the approximation (5.3) in the tensor product, total degree, hyperbolic cross, and Smolyak polynomial spaces, the required functions p and g are described in Table 5.1. Moreover, if the underlying abscissas are nested, as for the Clenshaw-Curtis points described above, the approximation (5.3) remains a Lagrange interpolant. For non-nested point families, such as standard Gaussian abscissas, the approximation (5.3) is no longer guaranteed to be an interpolant, but the analysis of the approximation error remains similar to the analysis presented here (see [31] for more details).*

5.2. Multilevel approximation using generalized sparse grid interpolants. The goal of the section is to verify the the assumptions of our multilevel collocation scheme for the generalized global sparse grid operator $\mathcal{I}_{M_k} = \mathcal{A}_{L_k}^{p,g}$. The convergence of the global sparse grid operators, applied to the the approximate solutions u_{h_k}, and the functionals $\psi(u_{h_k})$, is proved using the uniform ellipticity of the input coefficient in polyellipses containing Γ, based on complex analysis arguments. By $\Re(z)$ we denote the real and imaginary part of a complex number z, and require the additional assumption on the regularity of the coefficient a:

A7. (*Holomorphic parameter dependence*) The complex continuation of a, represented as the map $a^* : \mathbb{C}^N \to L^\infty(D)$, is a $L^\infty(D)$-valued holomorphic function on $\mathbb{C}^N$ that is $(\delta, \boldsymbol{\tau})$-polyellipse uniform elliptic, i.e., for all $\delta < a_{\min}$, there exists a vector $\boldsymbol{\tau} = (\tau_n)_{1\le n\le N}$ with $\tau_n > 1$ for all n, such that $\Re(a(x, \boldsymbol{z})) \ge \delta$ for all $x \in \overline{D}$ and all $\boldsymbol{z} = (z_n)_{1\le n\le N}$ contained in the polyellipse

$$\mathcal{E}_{\boldsymbol{\tau}} = \bigotimes_{1\le n\le N} \left\{ \frac{1}{2}\left(z_n + z_n^{-1}\right) : z_n \in \mathbb{C}, |z_n| = \tau_n \right\}.$$

The set $\mathcal{E}_{\boldsymbol{\tau}} \subset \mathbb{C}^N$ is the product of ellipses in the complex plane, with foci $z_n = \pm 1$, which are the endpoints of the domain $\Gamma_n, n = 1, \ldots, N$. Such polyellipses are commonly used in proving convergence results for global polynomial and interpolation schemes. In particular, the following result is proved in [13, Theorem 1.2] and [16, Lemma 3.3 and Theorem 2.5].

Lemma 5.4. (Analyticity of the solution u to (2.1)) *Assume that the coefficient $a(x, \boldsymbol{y})$ satisfies assumptions* **A3** *and* **A7** *for some $0 < \delta < a_{\min}$ and $\boldsymbol{\tau} = (\tau_n)_{1\le n\le N}$ with $\tau_n > 1 \ \forall n$. Then the function $\boldsymbol{z} \mapsto u(\boldsymbol{z})$ is holomorphic in an open neighborhood of the the polyellipse $\mathcal{E}_{\boldsymbol{\tau}}$.*

We remark that if the solution u admits less regularity, we might use local basis functions such as wavelets or splines to construct the interpolant. For a solution u which admits an analytic extension, convergence with respect to the total number of collocation points for the tensor product, sparse isotropic, and anisotropic Smolyak approximations (see Table 5.1) was analyzed in [1, 31, 32]. In what follows, our goal is to prove the bounds on the interpolation error in the approximate solutions u_{h_k} and the functionals $\psi(u_{h_k})$, for $k \in \mathbb{N}_0$, and thus verify the convergence assumptions given in **A5** and **F2, F3**. Under the polyellipse analyticity assumptions and Lemma 5.4, we arrive at the following result, which is given in [31, 32].

Theorem 5.5. *Let W denote a general Banach space and let $v \in C^0(\Gamma; W)$ admit an analytic extension in the complex polyellipse $\mathcal{E}_{\boldsymbol{\tau}}$. Then, with $r = \min_{1\le n\le N} \tau_n$, there exist constants $C(N)$ and $\mu(r, N)$, depending on N, such that*

$$\|v - \mathcal{A}_L^{p,g} v\|_{L^2_\rho(\Gamma;W)} \le C(N)\, M_L^{-\mu(r,N)}\, \zeta(v),$$

where M_L is the number of points used by $\mathcal{A}_L^{p,g}$ and

$$\zeta(v) \equiv \max_{\boldsymbol{z}\in\mathcal{E}_{\boldsymbol{\tau}}} \|v(\boldsymbol{z})\|_W. \tag{5.4}$$

Note that Theorem 5.5 only gives an algebraic convergence rate for the collocation approximation in terms of the number of points, while for large levels L the collocation scheme is known to converge exponentially [31,32]. Since our method is constructed using a combination of low- and high-level interpolants, the assumption of algebraic rates in practice apply to the full range of operators $\mathcal{I}_{M_k}$ used in the multilevel construction.

Define the Banach space $\Lambda(\Gamma; H_0^1(D))$ consisting of all functions $v \in C^0(\Gamma; H_0^1(D))$ such that v admits an analytic extension in the region $\mathcal{E}_{\boldsymbol{\tau}}$. It follows from Lemma 5.4 that, under appropriate assumptions on a, we have $u \in \Lambda(\Gamma; H_0^1(D))$. Because the dependence on $\boldsymbol{y}$ is unchanged in the approximate solution u_{h_k}, it also follows that $u_{h_k} \in \Lambda(\Gamma; H_0^1(D))$ for all $k \in \mathbb{N}_0$, and hence also $u_{h_k} - u_{h_{k-1}} \in \Lambda(\Gamma; H_0^1(D))$ for all $k \in \mathbb{N}$.

Similar to **A4**, using the triangle inequality and the assumption that $h_k = \eta^{-k} h_0$, it follows from standard finite element convergence theory [5, 10] that with $W = H_0^1(D)$ and ζ as in (5.4), $\zeta(u_{h_k})$ can be bounded by a constant independent of k, whereas $\zeta(u_{h_k} - u_{h_{k-1}})$ can be bounded by a constant multiple of h_k^β for some $\beta > 0$. In general, the constants appearing in these estimates will depend on norms of a and f as well as on the mesh refinement parameter η. We can hence conclude that with $\mathcal{I}_{M_k} = \mathcal{A}_{L_k}^{p,g}$, assumption **A5** is satisfied for the interpolation schemes considered in Theorem 5.5. Therefore, for the numerical examples presented in Section 6, we utilize the sparse grid stochastic collocation as the interpolation scheme.

Remark 5.6. **Dimension-dependent convergence rate**. *The asymptotic rate of convergence $\mu = \mu(r, N)$ in general deteriorates with growing dimension N of the stochastic space. For example, we have $\mu = r/N$ in the tensor product case, and for Smolyak sparse grids this is improved to $\mu = r/\log(N)$. The use of sparse grid SC methods is hence only of interest for dimensions N for which $\mu \ge 1/2$ so that the error still converges faster than the corresponding Monte Carlo sampling error. The multilevel approximation presented in this paper suffers from the same deterioration of convergence rate, and roughly speaking, the MLSC method can improve on the multilevel Monte Carlo method only when standard SC performs better that standard Monte Carlo; see [12, Theorem 4.1].*

Remark 5.7. **Anisotropic sparse grid approximations**. *To define anisotropic Smolyak approximations, we introduce a weight vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$ into the definition of g to reflect the relative importance of each dimension when selecting points, e.g., the anisotropic sparse Smolyak space uses $p(l) = 2^{l-1} + 1$, $l > 1$ and $g(\boldsymbol{l}) = \sum_{n=1}^N \frac{\alpha_n}{\alpha_{min}}(l_n - 1)$. The weight α_n is related to the size of the largest polyellipse $\mathcal{E}_{\boldsymbol{\tau}}$ on which the map $u : \Gamma_n \to C^0(\prod_{j\ne n} \Gamma_n, W)$ can be analytically extended. These weights can be computed either* a priori *or* a posteriori*;*

see [32, section 2.2]. For an isotropic grid*, all the components of the weight vector* $\boldsymbol{\alpha}$ *are the same so that one has to take the worst case scenario, i.e., choose the components of* $\boldsymbol{\alpha}$ *to all equal to the minimum* α_{min}*.*

Now we verify the analyticity assumption in Theorem 5.5 also for the functionals $\psi(u)$. Because Lemma 5.4 already gives an analyticity result for u, we use the following result, which can be found in [36], about the composition of two functions on general normed vector spaces.

Theorem 5.8. *Let* X_1*,* X_2*, and* X_3 *denote normed vector spaces and let* $\theta : X_1 \to X_2$ *and* $\nu : X_2 \to X_3$ *be given. Suppose that* θ *is analytic on* X_1*,* ν *is analytic on* X_2 *and* $\theta(X_1) \subseteq X_2$*. Then the composition* $\nu \circ \theta : X_1 \to X_3$ *is analytic on* X_1*.*

Hence, if we can show that ψ is an analytic function of u, we can conclude that $\psi(u)$ is analytic on $\mathcal{E}_{\boldsymbol{\tau}}$. To this end, we need the notion of analyticity for functions defined on general normed vector spaces, which we will now briefly recall.

Given normed vector spaces X_1 and X_2 and an infinitely Fréchet differentiable function $\theta : X_1 \to X_2$, we can define a Taylor series expansion of θ at the point ξ in the following way [6]:

$$T_{\theta,\xi}(x) = \sum_{j=0}^{\infty} \frac{1}{j!}\, d^j\theta(\xi)(x-\xi)^j, \tag{5.5}$$

where $x, \xi \in X_1$, the notation $(x-\xi)^j$ denoting the j-tuple $(x-\xi, \ldots, x-\xi)$ and $d^j\theta(\xi)$ denoting the j-linear operator corresponding to the j-th Fréchet differential $D^j\theta(\xi)$. The function θ is then said to be *analytic* in a set $Z \subset X_1$ if, for every $z \in Z$, $T_{\theta,z}(x) = \theta(x)$ for all x in a neighbourhood $N_r(z) = \{x \in Z : \|x - z\|_{X_1} < r\}$, for some $r > 0$. The following result now immediately follows from Theorem 5.8.

Lemma 5.9. *Let the assumptions of Lemma 5.4 be satisfied. Suppose* ψ*, viewed as a mapping from* $H_0^1(D)$ *to* $\mathbb{R}$*, is analytic in the set* $\mathcal{E}(u) \subset H_0^1(D)$*, and* $u(\boldsymbol{z}; x) \in \mathcal{E}(u)$ *for all* $\boldsymbol{z} \in \mathcal{E}_{\boldsymbol{\tau}}$*. Then,* $\psi \circ u$*, viewed as a mapping from* Γ *to* $\mathbb{R}$*, admits an analytic extension to the set* $\mathcal{E}_{\boldsymbol{\tau}}$*.*

Together with Theorem 5.5, now with $W = \mathbb{R}$, it then follows from Lemma 5.9 that assumptions **F2** and **F3** in Proposition 4.6 are satisfied for the interpolation schemes considered in this section, provided the functional ψ is an analytic function of u. Note that the function ζ in Theorem 5.5 acts on $\psi(u)$ instead of u in this case, leading to optimal convergence rates in h of the stochastic interpolation error.

To finish the analysis, we give some examples of functionals that satisfy the assumptions of Lemma 5.9. We in particular make use of the following result on Taylor expansions [6].

Lemma 5.10. *Let* $\theta : X_1 \to X_2$*, for normed vector spaces* X_1 *and* X_2*, and let* $Z \subset X_1$*. If* $\|d^j\theta(z)\| \leq C^j j!$ *for all* $z \in Z$ *and some* $C < \infty$*, where* $\|\cdot\|$ *denotes the usual operator norm, then* θ *is analytic on* Z*. In particular,* θ *is analytic on* Z *if* $\|d^j f(z)\| = 0$ *for all* $z \in Z$ *and all* $j \geq j^*$*, for some* $j^* \in \mathbb{N}$*.*

Example 5.11. (*Bounded linear functionals*) In this case, for any $v, w \in H_0^1(D)$, we have

$$d\psi(v)(w) = \psi(w) \qquad \text{and} \qquad d^j\psi(v) \equiv 0 \quad \forall\, j \geq 2,$$

which implies that ψ is analytic on all of (complex-valued) $H_0^1(D)$. Examples of bounded linear functionals include point evaluations of the solution u in one spatial dimension and local averages of the solution u in some subdomain $D^* \subset D$, computed as $\frac{1}{|D^*|}\int_{D^*} u\mathrm{d}x$, in any spatial dimension.

Example 5.12. (*Higher order moments of bounded linear functionals*) As a generalization of the above example, consider the functional $\psi(v) = \phi(v)^q$, for some bounded linear functional ϕ on $H_0^1(D)$ and some $q \in \mathbb{N}$. For any $v \in H_0^1(D)$, the differentials of ψ are

$$\begin{aligned} d^j\psi(v)(w_1,\dots,w_j) &= \phi(v)^{q-j}\prod_{i=1}^{j}(q-i+1)\,\phi(w_i), && 1 \le j \le q, \\ d^j\psi(v) &\equiv 0, && j \ge q+1, \end{aligned}$$

from which it follows that ψ is analytic on all of $H_0^1(D)$.

Example 5.13. (*Spatial L^2-norm*) Consider the functional $\psi(v) = \int_D v^2 \mathrm{d}x = \|v\|^2_{L^2(D)}$. For any $v \in H_0^1(D)$, the differentials of ψ are

$$\begin{aligned} d\psi(v)(w_1) &= \lim_{\delta\to 0}\frac{\int_D (v+\delta w_1)^2 - \int_D v^2}{\delta} = \lim_{\delta\to 0}\frac{\int_D \delta v w_1 + \int_D \delta^2 w_1^2}{\delta} = 2\int_D v w_1, \\ d^2\psi(v)(w_1,w_2) &= \lim_{\delta\to 0}\frac{2\int_D (v+\delta w_2)w_1 - 2\int_D v w_1}{\delta} = 2\int_D w_2 w_1, \\ d^j\psi(v) &\equiv 0 \ \ \forall\, j \ge 2, \end{aligned}$$

which implies that ψ is analytic on the entire space $H_0^1(D)$. For the functional $\psi(v) = \|v\|_{L^2(D)}$, we use Theorem 5.8 and the analyticity of the square root function on $(0,\infty)$ to conclude that ψ is analytic on any subset $\mathcal{E}(u) \subseteq H_0^1(D)$ not containing 0.

The analysis in this example can easily be extended to the functionals $\|v\|_{H_0^1(D)}$ and $\|v\|^2_{H_0^1(D)}$.

6. Numerical Examples.

The aim of this section is to demonstrate numerically the significant reductions in computational cost possible with the use of the MLSC approach. As an example, consider the following boundary value problem on either $D = (0,1)$ or $D = (0,1)^2$:

$$\begin{cases} -\nabla\cdot(a(\boldsymbol{y},\mathbf{x})\nabla u(\boldsymbol{y},\mathbf{x})) &= 1 \quad \text{for } \mathbf{x} \in D \\ u(\boldsymbol{y},\mathbf{x}) &= 0 \quad \text{for } \mathbf{x} \in \partial D. \end{cases} \tag{6.1}$$

The coefficient a takes the form

$$a(\boldsymbol{y},\mathbf{x}) = 0.5 + \exp\left[\sum_{n=1}^{N}\sqrt{\lambda_n}\,b_n(\mathbf{x})\,y_n\right], \tag{6.2}$$

where $\{y_n\}_{n\in\mathbb{N}}$ is a sequence of independent, uniformly distributed random variables on [-1,1] and $\{\lambda_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ are the eigenvalues and eigenfunctions of the covariance operator with kernel function $C(x,x') = \exp[-\|\mathbf{x}-\mathbf{x}'\|_1]$. Explicit expressions for $\{\lambda_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ are computable [19]. In the case $D = (0,1)$, we have

$$\lambda_n^{1\mathrm{D}} = \frac{2}{w_n^2+1} \quad \text{and} \quad b_n^{1\mathrm{D}}(\mathbf{x}) = A_n(\sin(w_n\mathbf{x}) + w_n\cos(w_n\mathbf{x})) \quad \text{for all } n \in \mathbb{N},$$

where $\{w_n\}_{n\in\mathbb{N}}$ are the (real) solutions of the transcendental equation

$$\tan(w) = \frac{2\,w}{w^2-1}$$

and the constant A_n is chosen so that $\|b_n\|_{L^2(0,1)} = 1$. In two spatial dimensions, with $D = (0,1)^2$, the eigenpairs can be expressed as

$$\lambda_n^{\mathrm{2D}} = \lambda_{i_n}^{\mathrm{1D}} \lambda_{j_n}^{\mathrm{1D}} \quad \text{and} \quad b_n^{\mathrm{2D}} = b_{i_n}^{\mathrm{1D}} b_{j_n}^{\mathrm{1D}}$$

for some $i_n, j_n \in \mathbb{N}$. In both one and two spatial dimensions, the eigenvalues λ_n decay quadratically with respect to n [7].

Let $a^*(\boldsymbol{z}, \mathbf{x}) = 0.5 + \exp\left[\sum_{n=1}^{N} \sqrt{\lambda_n} b_n(\mathbf{x}) z_n\right]$ be the complex extension of a. Given a multiindex $\nu \in \mathbb{N}_0^N$, it is easy to see that the mixed partial derivatives of a^* satisfy

$$\partial_\nu a^*(\boldsymbol{z}, \mathbf{x}) := \frac{\partial^{|\nu|} a}{\partial^{\nu_1} z_1 \dots \partial^{\nu_N} z_N}(\boldsymbol{z}, \mathbf{x}) = a(\boldsymbol{z}, \mathbf{x}) \prod_{n=1}^{N} (\sqrt{\lambda_n} b_n(\mathbf{x}))^{\nu_n}.$$

Thus, given $\boldsymbol{z} \in \mathbb{C}^N$, the power series

$$a^*(\boldsymbol{z}', \mathbf{x}) = \sum_{\nu \in \mathbb{N}_0^N} \frac{\partial_\nu a^*(\boldsymbol{z}, \mathbf{x})}{\nu!} \prod_{n=1}^{N} (z_n' - z_n)^{\nu_n}$$

converges for all $\boldsymbol{z}' \in \mathbb{C}^N$ such that $|z_n' - z_n| < \frac{1}{\sqrt{\lambda_n} \|b_n(\mathbf{x})\|_{L^\infty(D)}}, n = 1, \dots, N$, and thus $a(\boldsymbol{z}, \mathbf{x})$ satisfies assumption **A7**.

For spatial discretization, we use continuous, piecewise-linear finite elements on uniform triangulations of D, starting with a mesh width of $h = 1/2$. As interpolation operators, we choose the (isotropic) sparse grid interpolation operator (5.3), using p and g given by the classic Smolyak approximation in Table 5.1, based on Clenshaw-Curtis abscissas; see Section 5.

The goal of the computations is to estimate the error in the expected value of a functional ψ of the solution of (6.1). For fair comparisons, all values of ε reported are relative accuracies, i.e., we have scaled the errors by the value of $\mathbb{E}[\psi(u)]$ itself. We consider two different settings: in Section 6.1, we consider problem (6.1) in two spatial dimensions with $N = 10$ random variables whereas, in Sections 6.2 and 6.3, we work in one spatial dimension with $N = 20$ random variables. Because the exact solution u is unavailable, the error in the expected value of $\psi(u)$ has to be estimated. In Sections 6.1 and 6.2, we compute the error with respect to an "overkilled" reference solution obtained using a fine mesh spacing h^* and high interpolation level L^*. However, because this is generally not feasible in practice, we show in Section 6.3 how the error can be estimated when the exact solution is not available and one cannot compute using a fine spatial mesh and high stochastic interpolation level.

The cost of the estimators is computed as discussed in Section 4.2 and Remark 4.5, with $\gamma = d$, i.e., by assuming the availability of an optimal linear solver. For simplicity we assume $C_c = 1$; if this is not the case, then all costs reported simply need to be multiplied by C_c. For non-optimal linear solvers for which $\gamma > d$, the savings possible with the multilevel approach will be even greater than demonstrated below.

6.1. $\mathbf{d = 2, N = 10}$. As the quantity of interest, we choose the average value of u in a neighborhood of the midpoint $(1/2, 1/2)$, computed as $\psi(u) = \frac{1}{|D^*|}\int_{D^*} u(\mathbf{x})\mathrm{d}\mathbf{x}$, where D^* denotes the union of the six elements adjacent to the node located at $(1/2, 1/2)$ of the uniform triangular mesh with mesh size $h = 1/256$.

We start by confirming, in Figure 6.1, the assumptions of Proposition 4.6. The reference values are computed with spatial mesh width $h^* = 1/256$ and stochastic interpolation level $L^* = 5$. The expected value of the quantity of interest is $\mathbb{E}[\mathcal{I}_{M_5}\psi(u_{1/256})] \approx 0.049$.

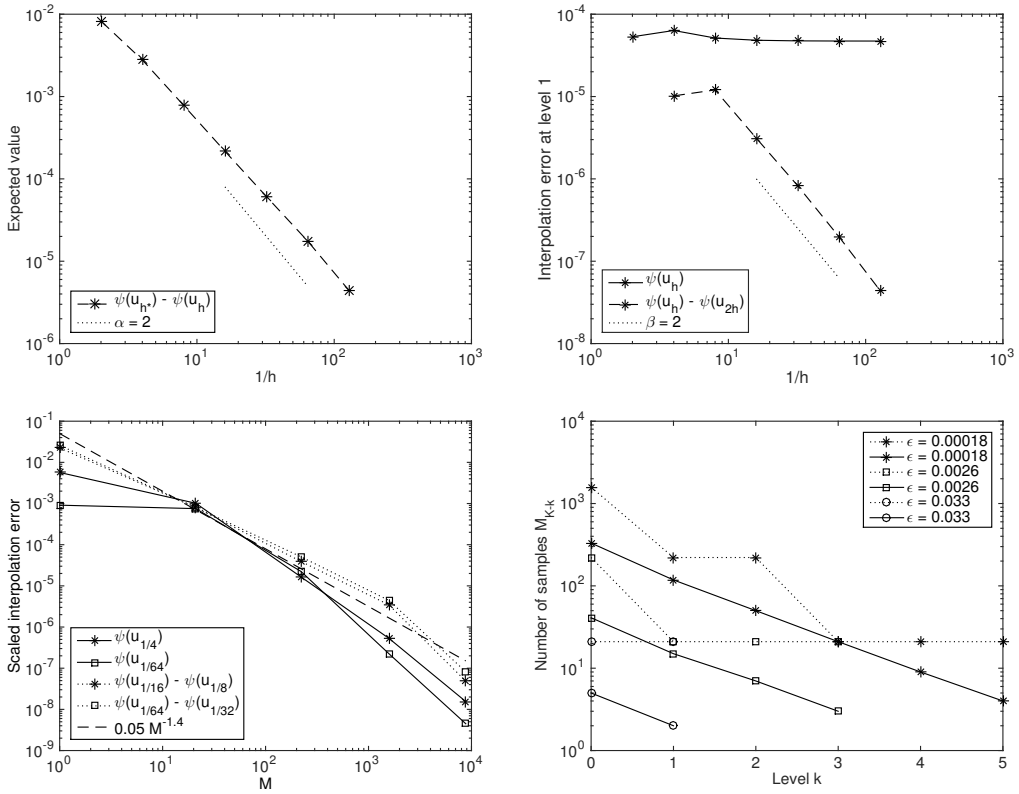

Figure 6.1. *$D = (0, 1)^2$, $N = 10$. Top left: $\mathbb{E}[\mathcal{I}_{M_5}\psi(u_h)]$ and $\mathbb{E}[\mathcal{I}_{M_5}(\psi(u_{1/256}) - \psi(u_h))]$ versus $1/h$ (assumption F1). Top right: $|\mathbb{E}[(\mathcal{I}_{M_5} - \mathcal{I}_{M_1})\psi(u_h)]|$ and $|\mathbb{E}[(\mathcal{I}_{M_5} - \mathcal{I}_{M_1})(\psi(u_h) - \psi(u_{2h}))]|$ versus $1/h$ (assumption F2*). Bottom left: $|\mathbb{E}[(\mathcal{I}_{M_5} - \mathcal{I}_{M_l})\psi(u_h)]/h_0^2|$ and $|\mathbb{E}[(\mathcal{I}_{M_5} - \mathcal{I}_{M_l})(\psi(u_h) - \psi(u_{2h}))]/h^2|$ versus M_l, for various h and $l = 0, \ldots, 4$ (assumption F2*). Bottom right: number of samples M_{K-k} versus k.*

The top-left plot of Figure 6.1 shows the convergence of the finite element error in the expected value of $\psi(u)$, and confirms that assumption **F1** of Proposition 4.6 holds with $\alpha = 2$.

Next, we verify assumptions **F2** and **F3**. To avoid the explicit computation of the function ζ from Theorem 5.5, we instead combine these assumptions into the assumption

$$\mathbf{F2^*}\text{:}\quad \left|\mathbb{E}\left[\psi(u_{h_k}) - \psi(u_{h_{k-1}}) - \mathcal{I}_{M_{K-k}}(\psi(u_{h_k}) - \psi(u_{h_{k-1}}))\right]\right| \leq C_I\, C_\zeta\, M_{K-k}^{-\mu}\, h_k^{\beta},$$

which is sufficient for the conclusions of Proposition 4.6 to hold.

The top-right plot of Figure 6.1 shows the absolute value of the interpolation error in the quantities $\psi(u_h)$ and $\psi(u_h) - \psi(u_{2h})$ for a fixed interpolation level $l = 1$, i.e. for fixed M_l,

as a function of h. We see that the interpolation error in $\psi(u_h)$ is bounded by a constant independent of h, whereas the interpolation error in $\psi(u_h) - \psi(u_{2h})$ decays quadratically in h. This confirms the convergence with respect to h in **F2***, with $\beta = 2$. Note that although only the interpolation error in $\psi(u_{h_0})$ is needed for the MLSC methods, the interpolation errors in $\psi(u_h)$, for general h, are needed to determine the number of samples in the standard SC approach.

The bottom-left plot of Figure 6.1 shows the interpolation error in $\psi(u_h)$ scaled by $h_0^2 = 4^{-2}$ and the interpolation error in $\psi(u_h) - \psi(u_{2h})$ scaled by h^2 for several values of h. According to assumption **F2***, these plots should all result in a straight line $CM^{-\mu}$, where $C = C_I C_\zeta$. The best fit which has $C = 0.05$ and $\mu = 1.4$ is added for comparison. Again, we have here added results for the functions $\psi(u_h)$, which are needed for the standard SC methods.

The bottom-right plot of Figure 6.1 shows the number of samples M_k computed using the formula (4.9), with $C = 0.05$ and $\mu = 1.4$, for several values of the relative accuracy ε. The finest level K was determined using the estimates on the finite element error from the top-left plot. Solid lines correspond to numbers rounded up to the nearest integer, as is done in (4.10), whereas dotted lines correspond to the number of samples rounded up to the next level of the sparse grid. Note that the cardinality of the sparse grids can be computed without actually computing the sparse grid. As stated in Remark 4.5, when the same number of points are used for consecutive levels, cancellations occur leading to savings in cost.

In Figure 6.2, we study the cost of the standard and multilevel collocation methods to achieve a given total accuracy ε. In both plots, the data labeled 'SC' and 'MLSC' denote standard and multilevel stochastic collocation, respectively. For the MLSC method, we consider two different approaches, labeled 'formula' and 'best', respectively. For both approaches, we choose $h_0 = 1/4$ and $\eta = 2$, and the finest level K is chosen such that the finite element error (with respect to the reference solution) is less than $\varepsilon/2$. For data labeled 'formula', the number of samples is then determined by the formula (4.9) with $C = 0.05$ and $\mu = 1.4$, rounded up to the next sparse grid level (the dotted lines in the bottom right plot of Figure 6.1). For data labeled 'best', the number of samples was chosen by trial and error so as to achieve a sampling error less than $\varepsilon/2$ for the smallest computational cost.

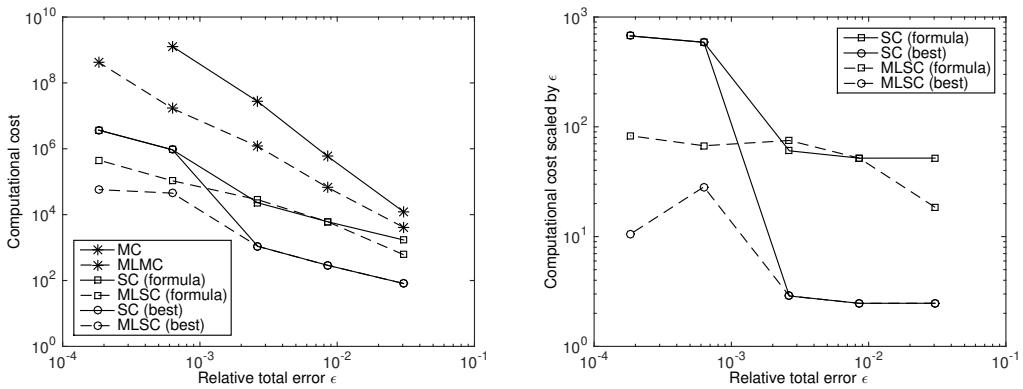

Figure 6.2. *$D = (0,1)^2$, $N = 10$. Left: computational cost versus relative error ε. Right: computational cost scaled by $\varepsilon^{-1.36}$ versus relative error ε.*

In the left plot of Figure 6.2, we simply plot the computational cost of the different estimators against ε. For comparison, we have also added corresponding results for Monte Carlo (MC) and multilevel Monte Carlo (MLMC) estimators. In both the 'formula' and the 'best' case, the multilevel collocation method outperforms standard SC. Both collocation-based methods outperform both Monte Carlo approaches. For the MLSC 'best' estimator, we observe that for the two smallest values of ε considered, the cost of the estimator only increases by a very small amount when decreasing ε, even though the finest level K is increased by one. This is due to the cancellations in computational cost discussed in remark 4.5.

In the right plot in Figure 6.2, we compare the computational cost shown in the left plot with the growth rates predicted by Proposition 4.6 for the standard and multilevel collocation methods. In our computations, we observed $\alpha \approx 2$, $\beta \approx 2$, and $\mu \approx 1.4$, which with $\gamma = 2$ gives computational costs of ε^{-1} and $\varepsilon^{-1.72}$ for the multilevel and standard SC method, respectively. We therefore plot the computational cost scaled by ε^1, expecting a constant line in the multilevel case, and a slope in the single level case. We see that both multilevel methods indeed seem to grow approximately like ε^{-1}, with the 'formula' case growing slightly faster for large value of ε and the 'best' case growing slightly faster for small values of ε. The costs for both standard collocation methods grow a lot faster with ε.

Figure 6.3 provides results for a different quantity of interest, $\psi(u) = \|u\|_{L^2(D)}$. The left plot corresponds to the bottom-left plot in Figure 6.1 and again confirms that the interpolation error in $\psi(u_h) - \psi(u_{2h})$ scales with h^2. The right plot corresponds to the left plot of Figure 6.2, where we plot the computational cost of the different estimators against ε. We see that all collocation-based methods outperform the Monte Carlo approaches. In both the 'formula' and the 'best' case, the multilevel collocation method again outperforms standard SC.

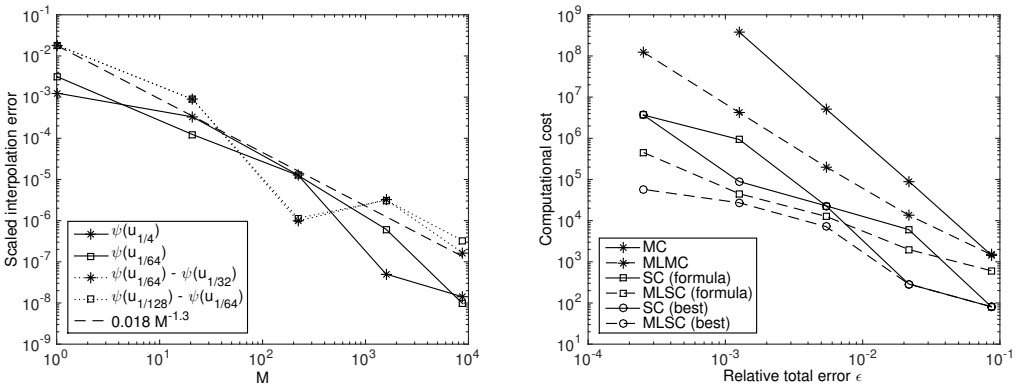

Figure 6.3. *$D = (0,1)^2$, $N = 10$. Left: $\mathbb{E}[\mathcal{I}_{M_5}\psi(u_h) - \mathcal{I}_{M_k}\psi(u_h)]/h_0^2$ and $[\mathcal{I}_{M_5}(\psi(u_h) - \psi(u_{2h})) - \mathcal{I}_{M_k}(\psi(u_h) - \psi(u_{2h})]/h^2$ versus M_k, for various h. Right: computational cost versus relative error ε.*

Remark 6.1. *Before considering the second model problem, let us briefly comment on the differences between the 'best' and the 'formula' multilevel methods. The 'formula' multilevel collocation method performs sub-optimally mainly for two reasons. First, it always rounds up the number of samples M_k to the nearest sparse grid level, which may be substantially higher than the number of samples actually required. Secondly, it does not take into account sign*

changes in the interpolation error, which in practice can lead to significant reductions in the interpolation error of the multilevel method. For both of these reasons, the interpolation error is often a lot smaller than the required $\varepsilon/2$, leading to sub-optimal performance. This issue is partly addressed in Section 6.3, where we consider not always rounding up, but rounding the number of samples either up or down to the nearest sparse grid level.

6.2. $\mathbf{d = 1, N = 20}$. We now repeat the numerical tests done in the previous section for the case $D = (0, 1)$ and $N = 20$. For the quantity of interest, we choose the expected value of the solution u evaluated at $x = \frac{3}{4}$. The reference values are computed using the mesh width $h^* = 1/1024$ and interpolation level $L^* = 5$. The expected value of the quantity of interest is $\mathbb{E}[\mathcal{I}_{M_5}\psi(u_{1/1024})] \approx 0.063$.

We again start by confirming, in Figure 6.4, the assumptions of Proposition 4.6. The four plots of that figure convey the same information as do the corresponding plots in Figure 6.1 and again confirm assumptions **F1**, **F2**, and **F3** of that theorem with $\alpha = 2$ and $\beta = 2$ and, in the bottom-left plot, the best line fit $C = C_I C_\zeta$ with $C = 0.005$ and $\mu = 0.8$.

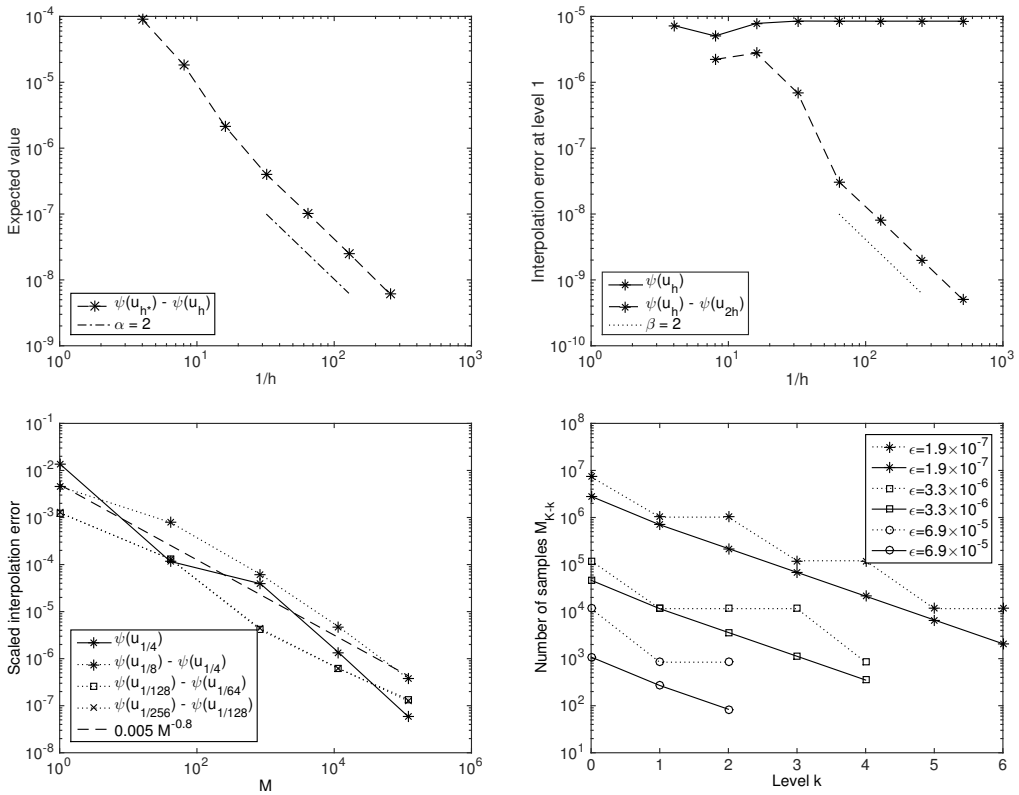

Figure 6.4. *$D = (0, 1)$, $N = 20$. Top left: $\mathbb{E}[\mathcal{I}_{M_5}\psi(u_h)]$ and $\mathbb{E}[\mathcal{I}_{M_5}(\psi(u_{1/1024}) - \psi(u_h))]$ versus $1/h$ (assumption F1). Top right: $|\mathbb{E}[(\mathcal{I}_{M_5} - \mathcal{I}_{M_1})\psi(u_h)]|$ and $|\mathbb{E}[(\mathcal{I}_{M_5} - \mathcal{I}_{M_1})(\psi(u_h) - \psi(u_{2h}))]|$ versus $1/h$ (assumption F2*). Bottom left: $|\mathbb{E}[(\mathcal{I}_{M_5} - \mathcal{I}_{M_l})\psi(u_h)]/h_0^2|$ and $|\mathbb{E}[(\mathcal{I}_{M_5} - \mathcal{I}_{M_l})(\psi(u_h) - \psi(u_{2h}))]/h^2|$ versus M_l, for various h and $l = 0, \ldots, 4$ (assumption F2*). Bottom right: number of samples M_{K-k} versus k.*

Figure 6.5 conveys the same information and uses the same labeling as does Figure 6.2.

Again, for both the 'formula' and 'best' cases, the multilevel collocation method eventually outperforms standard SC and both collocation-based methods also outperform the Monte Carlo approaches. Based on the values $\alpha \approx 2$, $\beta \approx 2$, and $\mu \approx 0.8$, Proposition 4.6 now predicts the computational costs of $\varepsilon^{-1.25}$ and $\varepsilon^{-1.75}$ for the multilevel and the standard collocation methods, respectively. The right-plot in Figure 6.5 indicates that the 'formula' multilevel collocation method indeed seems to grow like $\varepsilon^{-1.25}$ whereas the 'best' multilevel method actually seems to grow slower for small values of ε. For the three smallest values of ε considered, the computational cost is approximately the same, even though the finest level K is different in each case. This is due to the cancellations in computational cost discussed in Remark 4.5, as well as the different signs of the interpolation errors in the multilevel estimator. Also, again, the costs for both standard collocation methods grow a lot faster with ε.

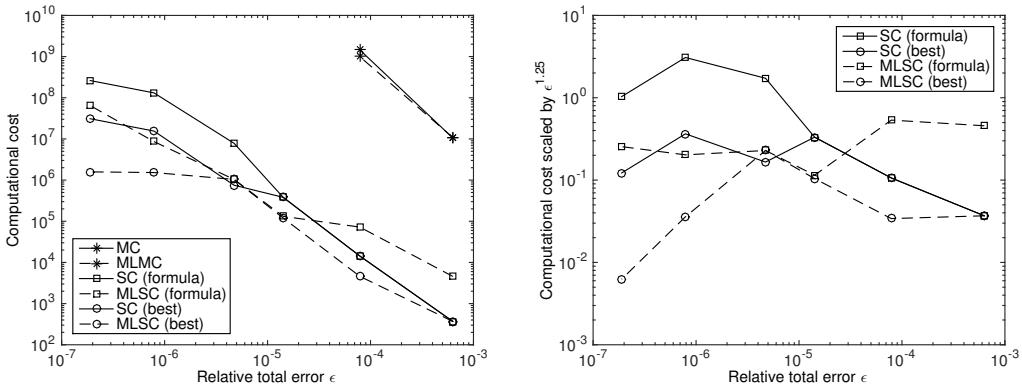

Figure 6.5. $D = (0,1)$, $N = 20$. *Left: computational cost versus relative error ε. Right: computational cost scaled by $\varepsilon^{-1.25}$ versus relative error ε.*

6.3. Practical implementation. In Sections 6.1 and 6.2, the accuracy of the computed estimates was assessed by comparison to a reference solution. Of course, in practice, a fine-grid, high-level reference solution is not available. Therefore, in this section, we describe how to implement the MLSC method without having recourse to a reference solution. We suggest the following practical strategy based on Richardson extrapolation, which is similar to the one proposed in [20]. In order to determine the number of levels we need, we assume that equality holds on assumption F1, i.e. we assume $\mathbb{E}[\psi(u) - \psi(u_{h_k})] = C_s h_k^\alpha$, and use the equality

$$\begin{aligned}\mathbb{E}[\psi(u_{h_k}) - \psi(u_{h_{k-1}})] &= \mathbb{E}[\psi(u) - \psi(u_{h_{k-1}})] - \mathbb{E}[\psi(u) - \psi(u_{h_k})] \\ &= C_s h_{k-1}^\alpha - C_s h_k^\alpha \\ &= (\eta^\alpha - 1)\,\mathbb{E}[(\psi(u) - \psi(u_{h_k}))],\end{aligned}$$

where we recall that $\eta = h_{k-1}/h_k$. Hence, the condition $\mathbb{E}[\psi(u) - \psi(u_{h_k})] \leq \varepsilon/2$ is equivalent to the condition $\mathbb{E}[\psi(u_{h_k}) - \psi(u_{h_{k-1}})] \leq (\eta^\alpha - 1)\varepsilon/2$. We then have the following algorithm.

1. Estimate the constants α, β, μ, and $C = C_I\, C_\zeta$.
2. Start with $K = 1$.
3. Calculate the optimal number of samples $M_k, k = 0, \ldots, K$, according to the formula (4.9), and round to the nearest sparse grid level.

4. Test for convergence by checking if there holds

$$\mathbb{E}[\psi(u_{h_K}) - \psi(u_{h_{K-1}})] \leq (\eta^\alpha - 1)\,\varepsilon/2.$$

5. If not converged, set $K = K + 1$ and return to step 3.

Note that in this procedure, steps 3 and 4 ensure that the interpolation error and the spatial discretization error are each less than the required tolerance $\varepsilon/2$, respectively.

The estimation of the constants α, β, μ, and C in step 1 can be done relatively cheaply from computations done using mesh widths h_0, h_1, and h_2 and interpolation levels $k = 0, 1, 2$. It is of course also possible to iterate over step 1, in the same manner as we iterate over steps 3 and 4, and to continuously update our estimates of these constants as we increase the number of levels in our multilevel estimator. This approach would eliminate some of the problems related to possible pre-asymptotic effects. It is also possible to use the idea behind the continuation MLMC (CMLMC) method in [14] and use a Bayesian approach to estimating the constants.

We test the algorithm using the the model problem from Section 6.2. For the results provided below, we estimated the convergence rate α from the level 1 interpolants $\mathcal{I}_{M_1}$ of $\psi(u_0)$, $\psi(u_1)$, and $\psi(u_2)$, resulting in $\alpha \approx 2.1$. In light of the results in Section 5, we assumed $\beta = \alpha$. We then used the first three interpolation levels of $\psi(u_0)$ and $\psi(u_1) - \psi(u_0)$ to obtain the estimates $C \approx 0.01$ and $\mu \approx 0.8$. Note that the value of μ is the same as in Section 6.2 whereas the value of the constant C is slightly larger. This is due to the fact that, for the large values of h used to estimate this constant, the function $\zeta(\psi(u_h) - \psi(u_{2h}))$ has probably not yet settled into its asymptotic quadratic decay.

As mentioned in Section 6.1, always rounding the number of samples resulting from formula (4.9) up to the next sparse grid level may lead to a substantial increase in the computational cost and hence a sub-optimal performance of the multilevel method. In practice, one might therefore consider not always rounding up, but instead rounding either up or down. As long as we do not round down more frequently than we round up, or at least not much more often, this approach should still result in an interpolation error below the required tolerance $\varepsilon/2$.

Table 6.1 shows the number of samples M_{K-k} resulting from the implementation described in this section for the model problem with $d = 1$ and $N = 20$ from Section 6.2. For each value of ε, the first row, denoted by 'formula', corresponds to the numbers M_{K-k} resulting from formula (4.9) rounded up to the nearest integer. The second row, denoted 'up', are the numbers in the first row rounded up to the next corresponding sparse grid level. For the final row, denoted 'up/down', the rounding of the number of samples was done in the following way: first, all numbers were rounded either up or down to the nearest corresponding sparse grid level. If this resulted in more numbers being rounded down than up, we chose the number that was rounded down by the largest amount and then instead rounded this number up. This procedure was continued iteratively. The same was done when more numbers were rounded up than down.

To confirm that the adaptive procedure still achieves the required tolerance on the total error, we have, for Table 6.2, computed the stochastic interpolation and finite element errors (with respect to a reference solution) and the computational cost of the multilevel approximations from Table 6.1. For comparison, we have added the results for the multilevel method

ε	level	0	1	2	3	4
6.3e-4	formula	191	48	15		
	up	841	841	41		
	up/down	841	41	41		
7.9e-5	formula	3002	747	233	73	
	up	11561	841	841	841	
	up/down	841	841	841	41	
1.4e-5	formula	27940	6949	2169	677	212
	up	120401	11561	11561	841	841
	up/down	11561	11561	841	841	841
4.7e-6	formula	110310	27433	8562	2672	834
	up	120401	120401	11561	11561	841
	up/down	120401	11561	11561	11561	841

Table 6.1

$D = (0, 1)$, $N = 20$. *Number of samples M_{K-k} computed using formula (4.9) and various rounding schemes.*

which was manually found to give a total error less than ε at minimal cost, which was already computed in Section 6.2 assuming a reference solutions was available. Note that for large values of ε, the adaptive procedure described in this section overestimated the finite element error, leading to a larger number of levels K compared to that found in Section 6.2. It is clear from Table 6.2 that not only does the alternative rounding procedure yield the required bound on the error, it also significantly reduces the computational cost of the multilevel method, bringing it close to what was manually found to be the minimal cost possible.

7. Concluding remarks. Computing solutions of stochastic partial differential equations using stochastic collocation methods can become prohibitively expensive as the dimension of the random parameter space increases. Drawing inspiration from recent work in multilevel Monte Carlo methods, this work proposed a multilevel stochastic collocation method, based on a hierarchy of spatial and stochastic approximations. A detailed computational cost analysis showed, in all cases, a sufficient improvement in costs compared to single-level methods. Furthermore, this work provided a framework for the analysis of a multilevel version of any method for SPDEs in which the spatial and stochastic degrees of freedom are decoupled.

The numerical results practically demonstrated this significant decrease in complexity versus single level methods for each of the problems considered. Likewise, the results for the model problem showed multilevel SC to be superior to multilevel MC even up to $N = 20$ dimensions.

One of the largest obstacles to the practicality of stochastic collocation methods is the huge growth in the number of points between grid levels. In the multilevel case, this can lead to a large amount of computational inefficiency. Certain simple rounding schemes were proposed to mitigate this effect, and proved to be extremely effective for the problems considered.

ε		Interpolation error	Spatial error	Cost
6.3e-4	up	6.7e-5	3.4e-5	8266
	up/down	2.8e-4	3.4e-5	4902
	best	8.0e-5	2.9e-4	369
7.9e-5	up	2.2e-5	6.3e-6	85558
	up/down	3.0e-5	6.3e-6	15650
	best	2.4e-5	3.4e-5	4591
1.4e-5	up	2.7e-6	1.6e-6	853207
	up/down	8.3e-6	1.6e-6	158714
	best	3.9e-6	6.3e-6	119699
4.7e-6	up	7.3e-8	1.6e-6	1519787
	up/down	1.2e-6	1.6e-6	1038183
	best	1.2e-6	1.6e-6	1038183

Table 6.2

$D = (0, 1)$, $N = 20$. Stochastic interpolation and spatial errors (with respect to the reference solution) and computational cost of various multilevel methods.

Similarly, since most of our example problems involved computation of a reference solution for the estimation of the necessary constants, a more practical multilevel stochastic collocation algorithm that dispensed with the need for a reference solution was proposed and tested.

It is clear that for any sampling method for SPDEs, whether Monte Carlo or stochastic collocation, multilevel methods are to be preferred over single-level methods for improved efficiency. Especially in the case of stochastic collocation methods, multilevel approaches enable one to further delay the curse of dimensionality, tempering the explosion of computational effort that results when the stochastic dimension increases. Though Monte Carlo methods are often preferable for problems involving a large stochastic dimension, multilevel approaches greatly improve the effectiveness of stochastic collocation methods versus Monte Carlo methods.

Acknowledgements. The first and the fourth authors are supported by the US Department of Defense Air Force Office of Scientific Research (AFOSR) under grant number FA9550-11-1-0149 and by he US Department of Energy Advance Simulation Computing Research (ASCR) program under grant number DE-SC0010678.

The second and third authors are supported by the US AFOSR under grant number 1854-V521-12 and by the US Department of Energy ASCR under grant number ERKJ259. Also supported by the Laboratory Directed Research and Development (LDRD) Program at the Oak Ridge National Laboratory (ORNL). The ORNL is operated by UT-Battelle, LLC, for the US Department of Energy under Contract DE-AC05-00OR22725.

REFERENCES

[1] I. Babuška, F. Nobile, and R. Tempone, *A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data*, SIAM J. Numer. Anal., 45 (2007), pp. 1005–1034.
[2] A. Barth, C. Schwab, and N. Zollinger, *Multi-level Monte Carlo Finite Element method for elliptic PDEs with stochastic coefficients*, Numer. Math., 119 (2011), pp. 123–161.
[3] J. Beck, F. Nobile, L. Tamellini, and R. Tempone, *Stochastic spectral galerkin and collocation methods for PDEs with random coefficients: a numerical comparison*, Lect. Notes Comp. Sci., 76 (2011), pp. 43–62.
[4] M. Bieri, *A sparse composite collocation finite element method for elliptic spdes.*, SIAM Journal on Numerical Analysis, 49 (2011), pp. 2277–2301.
[5] S. Brenner and L. Scott, *The Mathematical Theory of Finite Element Methods*, vol. 15 of Texts in Applied Mathematics, Springer, third ed., 2008.
[6] S. Chae, *Holomorphy and calculus in normed spaces*, Mg. Txb. Pu. Appl. Math., Marcel Dekker Inc, 1985.
[7] J. Charrier, *Strong and weak error estimates for the solutions of elliptic partial differential equations with random coefficients*, SIAM J. Numer. Anal., 50 (2012), pp. 216–246.
[8] J. Charrier, R. Scheichl, and A. Teckentrup, *Finite element error analysis of elliptic PDEs with random coefficients and its application to multilevel Monte Carlo methods*, SIAM J. Numer. Anal., 51 (2013), pp. 322–352.
[9] A. Chkifa, A. Cohen, and C. Schwab, *High-dimensional adaptive sparse polynomial interpolation and applications to parametric pdes*, Foundations of Computational Mathematics, 14 (2014), pp. 601–633.
[10] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North–Holland, 1978.
[11] C. W. Clenshaw and A. R. Curtis, *A method for numerical integration on an automatic computer*, Numer. Math., 2 (1960), pp. 197–205.
[12] K. Cliffe, M. Giles, R. Scheichl, and A. Teckentrup, *Multilevel monte carlo methods and applications to elliptic pdes with random coefficients*, Comput. Vis. Sci., 14 (2011), pp. 3–15.
[13] A. Cohen, R. Devore, and C. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic pde's*, Analysis and Applications, 9 (2011), pp. 11–47.
[14] N. Collier, A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R. Tempone, *A continuation multilevel monte carlo algorithm*, BIT Numerical Mathematics, (2014), pp. 1–34.
[15] S. Dereich and F. Heidenreich, *A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations*, Stoch. Proc. Appl., 121 (2011), pp. 1565–1587.
[16] R. Devore, H. Tran, C. Webster, and G. Zhang, *Analysis of quasi-optimal polynomial approximations for parameterized pdes with deterministic and stochastic coefficients*, (2015). To appear.
[17] P. Frauenfelder, C. Schwab, and R. Todor, *Finite elements for elliptic problems with stochastic coefficients*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 205–228.
[18] T. Gerstner and M. Griebel, *Dimension–adaptive tensor–product quadrature*, Computing, 71 (2003), pp. 65–87.
[19] R. G. Ghanem and P. D. Spanos, *Stochastic finite elements: a spectral approach*, Springer, New York, 1991.
[20] M. Giles, *Multilevel monte carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.
[21] M. Giles and C. Reisinger, *Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance*, SIFIN, 1 (2012), pp. 575–592.
[22] C. Gittelson, J. Könnö, C. Schwab, and R. Stenberg, *The multilevel Monte Carlo finite element method for a stochastic Brinkman problem*, Numer. Math., 125 (2013), pp. 347–386.
[23] M. Gunzburger, C. G. Webster, and G. Zhang, *An adaptive wavelet stochastic collocation method for irregular solutions of partial differential equations with random input data*, Springer Lecture Series on CS&E, (2013). To appear.
[24] M. Gunzburger, C. G. Webster, and G. Zhang, *Stochastic finite element methods for partial differential equations with random input data*, Acta Numer., 23 (2014).
[25] H. Harbrecht, M. Peters, and M. Siebenmorgen, *Multilevel accelerated quadrature for pdes with log-normal distributed random coefficient*, Preprint, 18 (2013).
[26] H. Harbrecht, M. Peters, and M. Siebenmorgen, *On multilevel quadrature for elliptic stochastic partial differential equations*, in Sparse Grids and Applications, Springer, 2013, pp. 161–179.
[27] S. Heinrich, *Multilevel monte carlo methods*, in Large-scale scientific computing, Springer, 2001, pp. 58–

67.

[28] H. Hoel, E. von Schwerin, A. Szepessy, and R. Tempone, *Adaptive multilevel Monte Carlo simulation*, in Numerical Analysis of Multiscale Computations, B. Engquist, O. Runborg, and Y.-H. R. Tsai, eds., vol. 82 of Lect. Notes Comp. Sci., Springer, 2012, pp. 217–234.

[29] X. Ma and N. Zabaras, *An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations*, J. Comput. Phys., 228 (2009), pp. 3084–3113.

[30] S. Mishra and C. Schwab, *Sparse tensor multi-level monte carlo finite volume methods for hyperbolic conservation laws with random initial data*, Mathematics of Computation, 81 (2012), pp. 1979–2018.

[31] F. Nobile, R. Tempone, and C. Webster, *A sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2309–2345.

[32] F. Nobile, R. Tempone, and C. G. Webster, *An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data*, SIAM J. Numer. Anal., 46 (2008), pp. 2411–2442.

[33] S. Smolyak, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, Dokl. Akad. Nauk SSSR, 4 (1963), pp. 240–243.

[34] A. L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann, *Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients*, Numer. Math., 3 (2013), pp. 569–600.

[35] H. van Wyk, *Multilevel sparse grid methods for elliptic partial differential equations with random coefficients*, arXiv:1404.0963, (2014). Submitted.

[36] E. Whittlesey, *Analytic functions in Banach spaces*, Proc. Amer. Math. Soc, 16 (1965), pp. 1077–1083.